

КОМПЮТЪРНИ НАУКИ И КОМУНИКАЦИИ



Електронно списание на
Център по информатика и технически науки на
Бургаски Свободен Университет

ТОМ 6, БРОЙ 1 (2017)

ISSN: 1314-7846

Списание „Компютърни науки и комуникации”,

Главен редактор

[prof. d-r Angel Toshkov](#), Bulgaria Burgas, 8001 SanStefano str. 62 Burgas Free University, Bulgaria

Редакционна колегия

1. [prof. d-r Andon Lazarov](#), Burgas Free University, Bulgaria
2. [prof. d-r Acad. Ivan Popchev](#), Bulgarian Academy of Sciences, Sofia, Bulgaria
3. [prof. d-r Mihail Iliev](#), Technical University of Ruse, Ruse, Bulgaria
4. [prof. d-r Kiril Konov](#), Technical University of Sofia, Research Institute of Communications, Sofia, Bulgaria, Bulgaria
5. [prof. d-r Ivan Kolev](#), Technical university of Gabrovo, Gabrovo, Bulgaria
6. [prof. d-r Vencislav Valchev](#), Technical University of Varna, Bulgaria
7. [prof. d-r Daniela Orozova](#), Burgas Free University, Burgas, Bulgaria
8. [prof. d-r Михаил Фёдорович Меняев](#), Доктор педагогических наук, профессор, Академик Международной Академии Информатизации, Москва, Россия
9. [assoc.prof. d-r. Veselina Jecheva](#), 62 San Stefano Str., Burgas
10. [assoc.prof. d-r. Evgeniya Petrova Nikolova](#), 62 San Stefano Str., Burgas
11. [assoc. prof. d-r Marcin Paprzycki](#), Associate Professor in the Systems Research Institute of the Polish Academy of Sciences, Warsaw, Poland
12. [d-r Yingjie Yang](#), School of Computer Science and Informatics Centre for Computational Intelligence (CCI) and De Montfort University Interdisciplinary Group in Intelligent Transport Systems (DIGITS)
13. [Michela Tramonti](#), Università degli Studi "G. Marconi" Rome Area, Italy

ISSN: 1314-7846

СПИСАНИЕ „КОМПЮТЪРНИ НАУКИ И КОМУНИКАЦИИ”,
ТОМ 6, NO 1 (2017), БСУ, БУРГАС

ЯАНУАРИ – ДЕКЕМВРИ, 2017 Г.

СЪДЪРЖАНИЕ

1. СРАВНИТЕЛНО РАЗПОЗАВАНЕ НА ЛИЦА СЪС СЕЛЕКТИРАНИ РЕГИОНИ С PRINCIPAL COMPONENT ANALYSIS (PCA) И KERNEL PRINCIPAL COMPONENT ANALYSIS (KPCA) И ФИЛТРИ НА ГАБОР - Петя Петрова – стр. 3-7
2. ПРЕДИЗВИКАТЕЛСТВА ПРЕЗ НАВЛИЗАНЕТО НА ИНФОРМАЦИОННИТЕ ТЕХНОЛОГИИ В БАЛКАНСКИТЕ УНИВЕРСИТЕТИ – КОНЦЕПТУАЛНИ ВЪПРОСИ И ПРАКТИЧЕСКИ ПРЕДИЗВИКАТЕЛСТВА, Божидар Божинов - стр. 8-19
3. ГЛАСОВО РАЗПОЗНАВАНЕ - ИСТОРИЧЕСКО РАЗВИТИЕ И ОСНОВНИ ТЕХНИКИ, Хасан Хасанов, Пенка В. Георгиева- стр. 20- 55
4. ФРАКТАЛНИ МОДЕЛИ В БИОЛОГИЯТА , Стефани Николова, Пенка В. Георгиева - стр. 56-68
5. РАЗПОЗНАВАНЕ НА ЛИЦА ОТ ВИДЕО ПОТОК В РЕАЛНО ВРЕМЕ, Ивелин Иванов, Виолета Вантева, Пенка В. Георгиева - стр. 69-98
6. ОРГАНИЗАЦИЯ УЧЕБНО-ТРЕНИРОВОЧНОГО ПОЛИГОНА ПРОТИВОКОРРОЗИОННОЙ ЗАЩИТЫ СЕЛЬСКОХОЗЯЙСТВЕННОЙ ТЕХНИКИ, Доцент к.воен.н. Алимов В.В - стр. 99-103

A COMPARATIVE STUDY OF FACE RECOGNITION BASED ON SELECTED REGIONS WITH PRINCIPAL COMPONENT ANALYSIS (PCA) AND KERNEL PRINCIPAL COMPONENT ANALYSIS (KPCA) AND GABOR FILTERS

Petya Petrova- Technical University- Varna, p5ia@abv.bg

Abstract: The paper presents a comparative study of performance for face recognition algorithms using Principal Component Analysis (PCA) and Kernel Principal Component analysis (KPCA). Images with various Regions of Interest (ROI's) are chosen from the databases to recognise faces. The results of parallel recognition are compared with results of ideal conditions. It has been established that the size of the ROI's affects the rate of recognition.

Keywords: Comparative Face Recognition, Region of Interest, Recognition Rate

СРАВНИТЕЛНО РАЗПОЗНАВАНЕ НА ЛИЦА СЪС СЕЛЕКТИРАНИ РЕГИОНИ С PRINCIPAL COMPONENT ANALYSIS (PCA) И KERNEL PRINCIPAL COMPONENT ANALYSIS (KPCA) И ФИЛТРИ НА ГАБОР

Петя Петрова- Технически университет- Варна, p5ia@abv.bg

Абстракт: В статията се разпознават лица чрез техниките Principal Component Analysis (PCA) и Kernel Principal Component analysis (KPCA) с използване на филтрите на Габор при не идеални условия. С двата метода паралелно се селектират различни по големина региони на интерес- Regions of Interest (ROI's) от всяко изображение в базата данни и се разпознават лица. Резултатите от паралелното разпознаване се сравняват с тези от разпознаването при идеални условия. Установено е как размерът на участъкът влияе върху процентът на разпознаване.

Ключови думи: сравнително разпознаване на лица, Регион на интерес (ROI), процент на разпознаване

Анотация. В статията се разпознават лица чрез техниките Principal Component Analysis (PCA) и Kernel Principal Component analysis (KPCA) с използване на филтрите на Габор при не идеални условия. С двата метода паралелно се селектират различни региони на интерес- Regions of Interest (ROI's) от всяко изображение в базата данни и се разпознават лица. Резултатите от паралелното разпознаване се сравняват с тези от разпознаването при идеални условия. Установено е как участъкът влияе върху процентът на разпознаване.

Ключови думи: сравнително разпознаване на лица, Регион на интерес (ROI), процент на разпознаване

1. Въведение

Разпознаването на лица се е превърнало в една от най-важните биометрични технологии. Нейни потенциални приложения са личната идентификация, контрол, сигурност, системи за наблюдение, телекомуникации и др. Пет са най-основните фактори, които могат да повлияят значително върху производителността на системата за разпознаване: осветление, поза, изражение, закрити участъци и застаряване [1].

В статията се ползват техниките за разпознаване на лица- Principal Component Analysis (PCA) и Kernel Principal Component analysis (KPCA) с използване на филтрите на Габор [1][2][3]. Разпознават се лица при не идеални условия. Селектира се регион(и) на интерес (Region(s) of Interest- (ROI's)) от всяко изображение в базата данни.

Установено е влияние на размерите на ROI's върху процентите от паралелното разпознаване.

1.2. Principal Component Analysis (PCA)

PCA се използва за намиране на t -измерно подпространство, чиито вектор съответства на максималната вариация (отклонение), където ($t < v$). Основните вектори се дефинират като собствени вектори *eigenvectors* от матрицата на разсейване и се дефинира като [4]:

$$\sum_{i=1}^M (x_i - \mu) \cdot (x_i - \mu)^T \quad (1)$$

Където μ означава стойност от всички M - изображения в обучаващия набор или дадено лице, T е транспозиция от неговите свойства и $|x_i|$ е i -тото изображение с неговите съединени колони във вектор. Основните компоненти от t - собствени вектори са с t - големи собствени стойности [4].

1.3. Kernel Principal Component analysis (KPCA)

Като се взема в предвид набора от изображения X_k [4],

$$x_k = [x_{k_1}, \dots, x_{k_n}]^T \in R^n \quad (2)$$

Kernel PCA проекциите на всеки вектор x от входното пространство, R^n до високо измерното пространство от характеристики R_f , чрез не линейно форматираната функция: $\Phi: R^n \rightarrow R^f$, $f > n$. След това PCA процесът се извършва въз основа на Kernel-подпространство чрез решаване на съответстващ проблем със собствена стойност:

$$\lambda \omega^\Phi = C^\Phi \omega^\Phi \quad (3)$$

Където C^Φ е ковариансна матрица. Всяко решение ω^Φ от $\lambda \neq 0$ принадлежи на участъка $\Phi(x_1), \dots, \Phi(x_m)$ [4].

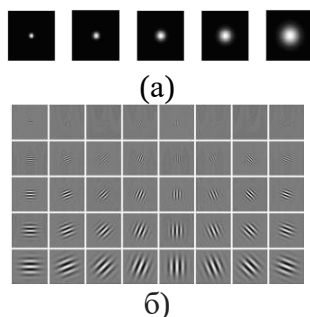
1.4. Представяне на Габор Уейвлети- 2D Gabor Wavelets (GW)

След като локалната честота и ориентация са не известни от тип $U \times V$ обикновено се предполага, че те извличат характеристики от изображението [4]:

$$\left\{ \varphi_{\Pi(f_u, \theta_v, \gamma, \eta)}(x, y) \right\}, f_u = \frac{f_{\max}}{\sqrt{2^u}}, \quad (4)$$

$$\theta_v = \frac{v}{8} \pi, +u = 0, \dots, U-1, v = 0, \dots, V-1$$

Където f_u и θ_v дефинират ориентацията и мащаба на Габор уейвлети, $f_{\max}$ е максималната честота, $\sqrt{2}$ и (половин октава) е фактор на отстояние между различните централни честоти. Стандартният Габор уейвлет има пет мащаба $U = 5$ и осем ориентации $V = 8$ за целият набор от 40 GW. На фиг. 1 са показани а) амплитуда от пет мащаба, на фиг. 1 б) реална част от пет мащаба и осем ориентации [5][6].



Фиг.1 Габор уейвлети

2. Изложение

Като предизвикателство пред редица изследователи [7] е въпросът, при който разпознаването на лица се извършва при липсващи части от лицето. За това те насочват

изследванията си, когато разпознаването се извършва посредством регион- ROI [8][9]. Разпознаването се затруднява- наличие на очила, шалове, позиция на лицето, изражения и др. [10], или липса на част лицето [10] (наличие на ROI). Холистичните методи (Eigenfaces и Fisherfaces), които извличат глобални характеристики в този случай са не приложими [10].

Настоящата статия има за цел сравняване на резултати от разпознаване на лица с холистичните методи PCA и KPCA и използване Габор уейвлети [2] при наличие на три различни размери ROI's. Стъпките по долу описват експериментът:

- Зареждане на изображения и изчисляване на амплитудата на Gabor-характеристиките
- Конструирание на филтърна банка на Габор (8X5) [2].
- Филтриране на изображението. Конструирание на дължината му, зареждане и извличане на характеристики от създадената матрица база данни (X) [2].
- Избиране на ROI's (x) ($x = X$ [Големина на участък])
- Разделяне на данните- обучаващ, оценяващ и тестови набори [2].
- Конструирание съответно на PCA/ KPCA - подпространство. Стартиране на оценяваща и тестова проекция на лице, извличане на характеристики [2].
- Изчисляване сходството на матрицата [2].
- Оценяване сходството на матрицата [2].

След изчисление на матрицата на сходство се визуализират и някои специфични, графични и математични резултати- ROC, СМС и ЕРС криви- показатели за ефективност [2].

На фиг. 2 е дадена блокова схема за тестване на процеса.



Фиг. 2 Блокова схема за тестване на процеса

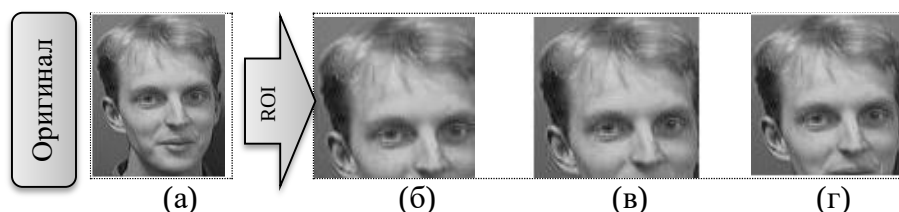
База данни е Olivetti Research Laboratory (ORL)- AT&T, 'pgm'. Всеки от 40-те субекта има по 10 различни изражения на лицето с включени промяна на осветлението. Това допълнително усложнява разпознаването [2]. Използва се Neighbor- класификатор, който да класифицира данните със сходство Mahalanobis Cosine (MANCOS)[2]. Данните се разделят като се образуват обучаващ, оценяващ и тестови набор от данни [2]

2.1.Разпознаване при идеални условия

Разпознаването при идеални условия с Габор филтри и PCA- без селектиране на ROI's е от порядъка на 74.17% [2]. А разпознаване с Габор филтри и KPCA в идеални условия е 80.00 % (отразено е в Таблица 1).

2.2.Регион(и) на интерес (Region(s) of Interest- (ROI's))

ROI са не идеално отсечените региони от всяко лице в базата данни. Селектират се три различни ROI's , с различни (близки) координати (размери) от всяко изображение на базата данни. С всеки от тях се формират три бази данни и се разпознават паралелно лица посредством съответно PCA/ KPCA и 40-те филтри на Габор. На фиг. 3 (а) е показано оригинално изоражение и условно обозначените малък фиг. 3 (б), среден фиг. 3 (в) и голям фиг. 3 (г) по размери участъци.



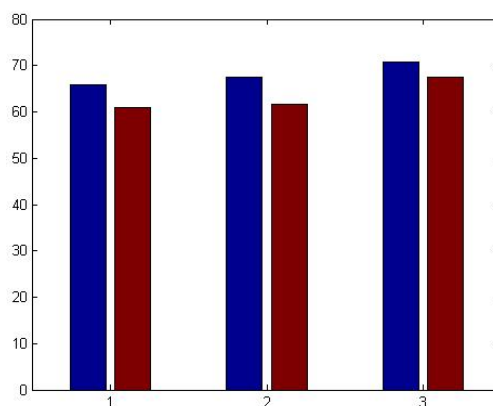
Фиг. 3 Различни по големина ROI's

В зависимост от големината на ROI се получават и различни резултати в разпознаването (Таблица 1), които са сравнени са с тези, получени в идеални условия (съответно PCA/KPCA+Габор)[2].

ROI's	Размери на ROI's	Разпознаване [в %] Габор	
		PCA	KPCA
	Малък	65.83	60.83
	Среден	67.50	61.67
	Голям	70.83	67.50
PCA+Габор- идеални условия	-	74.17	80.00

Таблица 1

Резултатите сочат това, че с увеличаване на ROI като размери, се увеличава и разпознаването [в %] (за двата метода). При най- малкият ROI е най-малко разпознаването. На фиг.4 са показани резултатите съответно за малък, среден, голям участък и за съответните методи PCA и KPCA и филтри на Габор.



1. ■ Габор+PCA
2. ■ Габор+KPCA

Фиг. 4 Резултати от разпознаването

3.Изводи

В статията е извършено разпознаване с PCA и KPCA и филтри на Габор при селектирани три различни участъка. С увеличаване размерите на ROI's, се увеличава и разпознаването, и обратно. В идеални условия разпознаване с KPCA генерира по-добър резултат (80.00 %) от този с PCA (74.17 %). Но при сравнение на разпознаването според съответните ROI's по-високи проценти се получават с PCA. Полученото максимално разпознаване при най-големият селектиран участък с PCA е 70.83 %, най-близко до разпознаване в идеални условия (74.17 %).

References

- [1] R. Ibrahim, Prof. F. Chadi, Prof. A. Samra. Plastic Surgery Face Recognition: A comparative Study of Performance. IJCSI International Journal of Computer Science Issues, ISSN : 1694-0814, Vol. 10, Issue 5, No 2, pp. 88-96, September 2013
- [2] V. Struc. The PhD Face Recognition Toolbox. Ljubljana, 2012
- [3] Isra'a Abdul-Ameer Abdul-Jabbar. Face Image Enhancement using Wavelet Denoising and Gabor Filters [J]. Department of Computer Science, University of Technology, IJCCCE Vol.16, No.1, 2016, Baghdad –Iraq, pp.104-117
- [4] Araoluwa S., Adetunmbi Adebayo O. Development of an Efficient Face Recognition System based on Linear and Nonlinear Algorithms, pp. 27-30
- [5] L. Shen and L. Bai. A review of Gabor wavelets for face recognition. *Patt. Anal. Appl.* 9, pp. 273-292, 2006
- [6] L. Shen, Li Bai. Information theory for Gabor feature selection for face recognition. *EURASIP Journal on Applied Signal Processing*, Vol. 2006, pp. 1–11
- [7] M. Oravec, J. Mazanec, J. Pavlovicova, P. Eiben, F. Lehocki. Face Recognition. Face Recognition in Ideal and Noisy Conditions Using Support Vector Machines, PCA and LDA. *Computer and Information Science*. ISBN 978-953-307-060-5
- [8] A. Goyal, N. Singh, M. Sood. Texture based Classification for Face Recognition. *Proceedings of SARC-IRF International Conference*, New Delhi, India, ISBN: 978-93-84209-08-7, pp. 50-54, 2014.
- [9] D. Vukadinovic, M. Pantic. Fully Automatic Facial Feature Point Detection Using Gabor Feature Based Boosted Classifiers. *IEEE International Conference on Systems, Man and Cybernetics Waikoloa*, pp.1-10, pp.1693-1698, Hawaii October, 2005
- [10] A. Andres, S. Padovani, M. Tepper, J. Berles. Face Recognition on Partially Occluded Images Using Compressed Sensing. *Pattern Recognition Letters*, 2013

CHALLENGES FOR THE ENTRY OF INFORMATION TECHNOLOGIES IN THE BALKAN UNIVERSITIES - CONCEPTUAL ISSUES AND PRACTICAL CHALLENGES

Bojidar Bojinov, Tsenov Academy of Economics, b.bojinov@uni-svishtov.bg

Abstract: This article aims to systematize and define the main challenges for the management of information technology in modern universities. In this aspect stands out research and thesis that the management of information technology in higher education institutions differ by a number of unique specificity than in classical companies, which are based on the unique combination of teaching, research and administrative activities within the university. The article is structured in two relatively distinct parts, at first do historical analysis and overview of the different aspects and stages of penetration of information technology in universities. The second part focuses on the real problems faced by universities from Balkan countries and presents the results of the first author of the international study "Challenges to the management of information technology in universities" (31.10-30.11.2016).

Keywords: information technology, university, strategic governance, educational innovation

JEL: m25

ПРЕДИЗВИКАТЕЛСТВА ПРЕЗ НАВЛИЗАНЕТО НА ИНФОРМАЦИОННИТЕ ТЕХНОЛОГИИ В БАЛКАНСКИТЕ УНИВЕРСИТЕТИ – КОНЦЕПТУАЛНИ ВЪПРОСИ И ПРАКТИЧЕСКИ ПРЕДИЗВИКАТЕЛСТВА

Божидар Божинов, Стопанска академия „Д.А. Ценов“, b.bojinov@uni-svishtov.bg

Резюме: Настоящата статия си поставя за цел да систематизира и дефинира основните предизвикателства пред управлението на информационните технологии в съвременните университети. В този аспект изпъква и изследователската теза, че управлението на информационните технологии във висшите училища се различава с редица уникални специфики от това в класическите фирми, които са породени от уникалното съчетание на преподавателска, научно-изследователска и административна дейности в рамките на университета. Статията е структурирана в две относително обособени части, като в първата се прави исторически анализ и преглед на различните аспекти и етапи на навлизането на информационните технологии в университетите. Втората част акцентира върху реалните проблеми, пред които са изправени висшите училища от балканските държави и представя резултатите от осъщественото от автора първо международно изследване „Предизвикателства пред управлението на информационните технологии в университетите“ (31.10-30.11.2016 г.).

Ключови думи: информационни технологии, университети, стратегическо управление, образователни иновации

JEL: m25

Възникването и развитието на съвременните информационни технологии е неразривно свързано с **университета като създател на знания**.¹ Трудно можем да открием еднозначно приноса на отделните учени, но сред ключовите открития, довели до появата на информационните технологии, такива каквито ние ги познаваме, можем да открием изобретяването на бройни пръчки на Непер, суматора на Да Винчи, „сметачния часовник“ на Шикард, „сумиращата машина“ на Паскал, „степенния числител“ на Лайбниц, и аритмометрите на Томас и Однер, изследванията на Бабидж, Лъвлейс, Бул, Пиърс, Шенън, Стибиц и редица други (Казакова, 2011). Сред разработките, довели до появата на предшествениците на съвременните компютри можем да открием тези на Цузе (Z-1), Стибиц (Bell-I), Айкен (Mark-1), Атанасов и Бери (ABC), Моучли и Екерт (ENIAC), както и ключовата роля на Агенция за напреднали изследователски проекти (ARPA)² и Калифорнийските университети в Лос Анджелис и Санта Барбара, Станфордския университет и Университета в Юта за появата на ARPANET³ (1969) – основа на съвременния Интернет. Зад всички тези открития и нововъведения стоят имена от университетските среди, чиито идеи са променили съвременния свят.

Втората исторически възникнала роля на университета в изследвания аспект е в качеството му на **ползвател на информационните технологии за улесняване на научните изследвания** чрез автоматизация на изчислителния процес. Първите създадени компютри, като част от изследователски проекти, са предназначени за ускоряване на изследванията чрез предоставяне на ефективен способ за намаляване времето за изчислителната работа на учените. По този начин, университетите се превръщат в своеобразни центрове на нововъзникващите технологии, а финансирането на тяхното развитие, често е свързано с предоставянето на ползването изчислителна мощ споделено в университета или срещу заплащане (практика, която се използва и до днес при ползването на големи изчислителни системи). Огромната изчислителна мощ на мейнфрейм системите позволява тестването на модели, създаването на симулации, търсенето на оптимизационни варианти и като цяло използването на сложен математически апарат във всички сфери на науката и изследванията, чието прилагане с традиционните изчислителни средства би отнело понякога повече време от човешкия живот.

Тратата базисна насока за навлизане на информационните технологии в дейността на университетите е свързана с ползването им за **автоматизация на административната им дейност** (Донева & Касъкчиев, 2006, с. 31). Този процес допълнително се засилва с разширяването на академичната автономия, особено в частта ѝ управление на активите и финансовите средства на университета, вкл. и тяхната отчетност. В рамките на този процес могат да се открият много общи черти и характеристики, присъщи за навлизането на информационните технологии и с дейността на останалите икономически субекти в обществото, и най-вече с намаляване на времето,

¹ Изследвайки тази роля, ние трябва да отбележим, че приведените факти и разсъждения, особено що се отнася до ранните периоди, са с известна степен на условност, тъй като от една страна не са оцелели достатъчно на брой надеждни и безспорни източници, и същевременно следва да отчитаме спецификата на създаване на науката. Тези доводи важат в пълна сила особено за Средновековието, където част от науката се е създавала в различни средища, а част от откритията са свързани със съвместна работа и писмена кореспонденция между водещи за времето си научни светили.

² Целта на агенцията е да финансира иновативни изследователски проекти, които могат да намерят приложение в отбраната, като участници и бенефициенти на финансирането са водещи американски университети (Wikipedia, 2014).

³ Първият директор на агенцията Ликлидър, си представя ARPANET като една „галактическа мрежа“ – глобална компютърна мрежа, която може да улесни обмяната на програми и данни. Това виждане за компютъра като устройство за подобряване на комуникацията е в рязък контраст с мисленето на производителите на компютри, които разглеждат компютъра като машина за изчисления. (Quinn, 2015, с. 28)

вложените усилия и броя на допуснатите човешки грешки в рутинните дейности на фирмата, най-често свързани с нейното счетоводно и финансово отчитане и други специфични дейности, изискващи голяма по обем и еднотипна изчислителна работа.

Тук обаче следва да отбележим известна разлика, дотолкова до колкото навлизането на компютърните технологии във фирмената дейност е свързано с придобиване на компютри и поддържането на специализиран екип за тяхната поддръжка и програмиране, докато при университетите, особено при технологичните, компютърната инфраструктура е на разположение, а използването ѝ за административни и бизнес дейности е свързано по-скоро с желанието за тяхното прилагане от отговорните за тези решения лица. Друга съществена разлика, която може да се открие, е, че ако за фирмите внедряването на нови информационни технологии е пряко свързано с повишаване на ефективността от основната им бизнес дейност, което се явява мощен стимул за този процес, то при университетите този стимул отсъства (Caspary, Kusserow, Lavin & Movassaghi, 1999, с. 5). Това от своя страна става причина за известно забавяне на технологичното проникване на новите информационни технологии в административната и бизнес дейността на университетите, процес, който започва да търпи промяна с навлизането на фирмените подходи и механизми в управлението на висшите училища.

Друга съществена пречка за имплементиране на водещите технологични решения и системи за фирмено управление в рамките на университетите е спецификата на образователния процес, като основна бизнес дейност за университетите. Трансферирането на понятия и концепции като „продукт“, „клиент“, „качество“ във вида им, приложим в производствената сфера, се оказва сериозно предизвикателство, ако не и неприложимо в областта на висшето образование. Това предопределя и частичната информатизация на университетите единствено в областта на техните бизнес-административни дейности, където допирните точки с реалния бизнес са значително повече. Едва през последните десетилетия започва да се говори за създаването на **образователни управленски системи**, които да автоматизират основната административна дейност на университетите.

През последните десетилетия можем да говорим и за проявата на нова и специфична насока за проникване на информационните технологии във висшите училища, свързана с тяхното **използване в педагогическата работа и процеса на обучение**, като нов способ за представяне на информация, средство за комуникация и начин за усвояване на нови знания. Този процес може да се разглежда като следствие от тенденцията за информатизация на обществото и изграждането на глобални информационни мрежи.

От една страна, новата информационно базирана икономика на знанието поставя нови изисквания пред университетското образование, по отношение притежаването на определено ниво на компютърна грамотност за завършващите, позволяваща им адекватно да се реализират на пазара на труда и да покрият належащите изисквания от страна на работодателите (Лазаров, 2016, с. 105). Това доведе до навлизането в учебния процес на дисциплини, свързани с общата работа с компютри, приложен софтуер и мрежи, съответстващо на нивото на компютъризация на реалния бизнес сектор.

От друга страна, навлизането на новите форми за аудио-визуално комуникиране, позволи на преподавателите да променят начина на представянето на учебния материал, като въведат интерактивен характер в обучителния процес. Извън традиционното използване на мултимедийни способи за презентиране, трайно в учебния процес навлязоха симулационните методи, системите за самообучение и самоконтрол на знанията. От студентите започна да се изисква не само възпроизвеждане на представения

материал, но и презентиране на самостоятелно извършени научни изследвания и проучвания, вкл. чрез използване на онлайн ресурси и библиотеки.

С масовизацията на информационните технологии, обусловено от драстичния спад на цените на хардуерните устройства, съчетано с повишаване на достъпността до ресурсите на информационните мрежи и Интернет, университетите започнаха за използват новите технологии за пазарна експанзия и за достигане до нови потенциални клиенти. Наличието на високоскоростна интернет връзка, наличието на системи за видео-конферентни връзки и появата на специализирани системи за електронно обучение, стимулираха развитието на нов сегмент от висшето образование – дистанционното обучение (Nicolet, 2011, с. 25-26). То се оказва особено атрактивно за всички участници в образователния процес, до колкото позволи на университетите да намалят цената на предлагания от тях образователен продукт, а на студентите – да получат същите по обем и качество знания, с възможност за съчетаване на образованието с тяхната професионална трудова реализация. Тази тенденция, особено след появата на т.нар. МООС курсове (Selingo, 2014, с. 5), дори води до промяна не само на съществуващите образователни модели, но и на базисните бизнес модели на университетите.

В изследването си Пенрод, Доленс и Дъглас обособяват висшите училища в три големи групи от гледна точка тяхното отношение и визия за използването на информационните технологии в дейността им (Penrod, Dolence & Douglas, 1990, с. 8):

- Институции, възприемащи информацията и информационните технологии като стратегически ресурс и основа за постигане на конкурентно предимство, повишаване на качеството и постигане на високи резултати;
- Институции, възприемащи информацията и информационните технологии като помощно средство за повишаване на ефективността на неосновни ежедневни дейности;
- Институции, с липса на ясна визия за ролята на информацията и информационните технологии във висшето образование.

В специализираната литература съществуват много изследвания, посветени на навлизането на информационните технологии във фирмената дейност, като акцентът в тях се поставя както върху процеса на информатизация на фирмата, така и върху използваните типове управленски ИТ системи (Ward & Peppard, 2002, с. 9.). С оглед визираната по-горе специфика на процесите на навлизане на информационните технологии в университетите, буквалното взимстване и адаптиране на тези модели е неприложимо.

Подобен опит за систематизация можем да открием в изследване на Санял, в което се акцентира върху навлизането на информационните технологии в административната дейност на университета, определяйки три базисни фази на този процес (Sanyal, 1995, с. 2). Според него, всяка фаза се характеризира със следните особености (Sanyal, 1995, с. 2-3):

- за **фаза 1 „Голяма система“** е специфично, че университетите използват централизирани мейнфрейм системи, основно за административни цели, като между изградените информационни системи липсва интеграция (с изключение на частично споделени файлове), и чиято цел е обработката на големи количества данни, предимно за нуждите на външни потребители;
- при **фаза 2 „Децентрализация“** се наблюдава преход към ведомствени компютри, чиято основна цел е обработка на информация за нуждите на средния академичен мениджмънт, като за целта се разширява и обхвата на дейностите и звената обслужвани от информационните системи;

- **фаза 3 „Рейнтеграция“** се характеризира със засилени интеграционни процеси между информационните системи и звената в университета, подобряване на електронния документооборот, и нарастване на значението на резултатната информация за управленския мениджмънт.

Интегрирайки констатациите в изследването на Суес (Suess, 2016), ние можем да обособим още два допълнителни етапа разширявайки и допълвайки класификацията на Санял, а именно:

- **фаза 4 „Интеграция на културите“** при която се създават системи и технологии, които да интегрират различните аспекти от университетската култура в единна информационна среда, позволяваща университета да получи нови конкурентни и стратегически предимства от интегрираното академично управление. При този подход бизнес и ИТ целите се поставят в услуга на постигането на стратегическите цели на институцията.
- **фаза 5 „Добавяне на стратегическа стойност“** за университета чрез редизайн на образователните, научните и бизнес процесите в институцията, с цел адаптиране към предизвикателствата на новата икономика на знанията и променящите се образователни и бизнес модели на функциониране на висшите училища.

Сред значимите изследвания посветени на този аспект от навлизането на информационните технологии в икономиката, и в частност – във висшето образование, можем да открием на Николет (Nicolet, 2011). Според него, в исторически аспект процеса може да се обособи в няколко периода:

- **Първи етап** (от края на 50-те до средата на 60-те години на XX век) – автоматизация на отделните дейности и операции, с акцент върху обработката и управление на административни данни на база използването на мейнфрейм компютри и перфокарти.⁴ Основните дейности, подлежащи на автоматизирана обработка са свързани с управление на счетоводство и финанси, както и отделни аспекти от управлението на персонала и административното състояние на студентите (Nicolet, 2011, с. 19).⁵ В свое изследване Риланд посочва, че въпреки, че компютрите се използват за научни изчисления и компютърно-базирани изследвания, тяхното основно приложение остава автоматизация на административните дейности (цитирано по Nicolet, 2011, с. 19).
- **Втори етап** (от средата на 60-те до края на 70-те години на XX век) – засилване на ролята на компютърните и информационни технологии в качеството на учебни и изследователски инструменти. Промяната настъпва в резултат на както на изследванията на Росър и Пиърс, така и на докладите на Комисията за ползване на компютри (САЩ) “Нуждата от цифрови компютри в университетите и колежаите” и на Консултативния комитет по науката към Президента на САЩ „Компютрите във висшето образование“, които довеждат до създаване на специализирана Служба за компютърни дейности към Националната научна фондация на САЩ. В резултат на целевото държавно финансиране само за 5 години броя на компютрите в американските

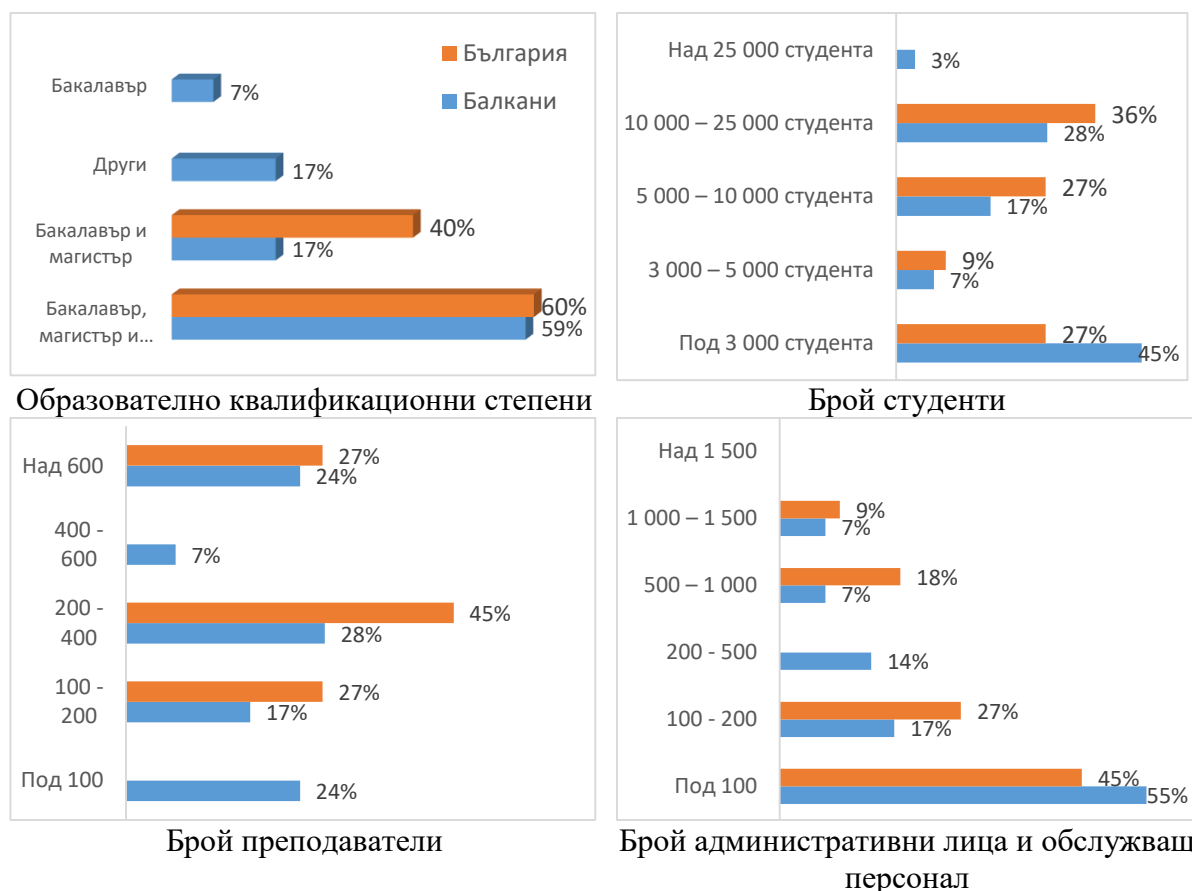
⁴ Само за сравнение можем да посочим, че в рамките на Стопанска академия посочената технология бе използвана почти до средата на 90-те години на XX век.

⁵ За значимостта на този процес говори и проведената през 1956 г. специализирана конференция, посветена на обработка на данните в университетите „CUMREC – A Higher Education Administrative Technology Conference“. Вж. Nicolet (2001, с. 19)

университети нараства с 20%, макар че тяхното основно научно предназначение е свързано основно с математиката и компютърните науки (цитирано по Nicolet, 2011).

- **Трети етап** (от началото на 80-те до средата на 90-те години на XX век) – навлизане на микро компютрите във всички дейности на университетския живот. Дразтичната технологична промяна, свързана с широкото навлизане на персоналните компютри, способствана и от многократно по-ниската им цена спрямо доминантните до тогава мейнфрейм системи, позволява компютъра да стане неразривна част от всяко университетско бюро. Многократно се улеснява достъпа до изчислителни и информационни ресурси, както за целите на управлението на административните процеси, така и във всички сфери на научните изследвания и процеса на преподаване и усвояване на знания. Разбира се, навлизането на новия вид компютърна техника не заменя използването на мощните компютърни системи – те все още се явяват ядрото на централизираната обработка на информацията, а персоналните компютри, които са със значително по-малка изчислителна мощ намират своето приложение за автоматизация на ежедневните офис дейности и замяна на терминалните станции.
- **Четвърти етап** (от средата на 90-те до края на 90-те години на XX век) – проникване на Интернет в университетската дейност. Макар, че университетите са създатели на първообраза на съвременния Интернет, процесите на неговата комерсиализация, съчетана с новите комуникационни способности (електронна поща, форуми, специализирани бази данни) и промяната в начина на визуализация на информацията (уеб сайтове) доведоха до своеобразна революция в използването на новите информационни технологии в университетите. Академичните общности осъзнаха потенциала за новите медии за разширяване на пазарния си дял и ускоряване масовизацията на висшето образование чрез различни способности за електронно и дистанционно обучение.
- **Пети етап** (първото десетилетие на XXI век) – дигитализация на учебния процес и разширяване на географския обхват на преподаване. Създаването на дигитални и мултимедийни ресурси се превърна в неотменна част от дейността на университетите. За обслужването на тези нови и специфични информационни нужди се появиха цели нови класове информационни системи (CMS, LMS), а иновативните методи за електронно преподаване навлязоха като пълноценен елемент допълващ класическите педагогически подходи и в редовните форми на обучение. Новите образователни и бизнес реалности наложиха висока степен на интеграция на университетските информационни системи, както и прилагането на водещи корпоративни практики за тяхното управление.
- **Шести етап** (второто десетилетие на XXI век) – навлизане на мобилните устройства, социалните мрежи и облачните технологии в дейността на университетите. Нововъзникващите технологии продължават да променят съвременното ни общество, като поставят и редица нови предизвикателства пред университетите. Повишената достъпност на изчислителна мощ и интернет свързаност чрез персоналните преносими устройства (телефони, таблети, лаптопи) променя и акцентите в използването на информационните технологии. Достъпността до образователни, научни и информационни ресурси, навсякъде и по всяко време е ключова за съвременното общество, като акцента от използване на системи за автоматизация се пренасочва към

създаване на знания и придобиване на стратегически предимства. Ученето от еднократен акт се променя в непрекъснат процес, продължаващ целия съзнателен живот на индивида, а ограниченото притежаване и придобиване на знания се превръща в отворена и безплатна образователна среда.



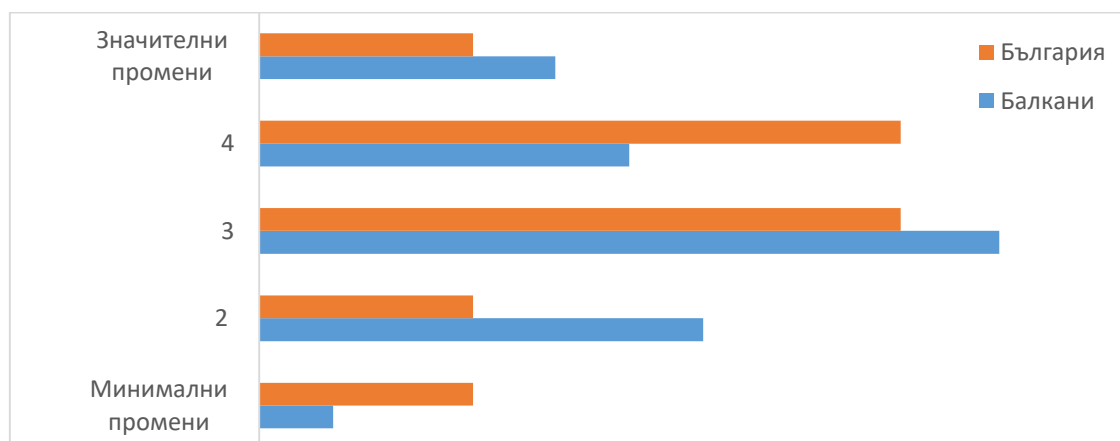
Фиг. 1. Разпределение на участващите университети по различни критерии

За да бъдат систематизирани и дефинирани **общите проблеми пред университетските общности в процеса на управление на информационните технологии във висшите училища**, в края на 2016 година бе проведено международно проучване сред университетите в Централна и Източна Европа. Обследвани са 246 висши училища, разположени в държави от Балканския полуостров, като сред тях преобладават тези с **държавно или обществено финансиране** (66% за анализирания извадка и 90% за България). Що се отнася до образователно-квалификационните степени по които се обучават студентите в тези университети, то сред участниците доминират висшите училища обучаващи във всички форми, предвидени от законодателя (59% за анализирания извадка и 60% за България), следвани от предоставящите обучение в ОКС „бакалавър“ и „магистър“ (вж. фиг. 1)

Изследването констатира **висока степен на развитие и разпространение на информационните технологии в балканските университети** (62%) и значително по-малка степен на нерационалност на очакванията от страна на студентите (19%) и обществеността (19%). Преобладаващата част от анкетиранияте (46% - Балкани, 60% - България) посочват, че **информационните нужди в университета са добре разбрани, като отговорностите по тяхното управление са насочени към централизирано ИТ звено**, докато значително по-малка част използват смесения (38% - Балкани, 20% -

България) или децентрализирания (15% - Балкани, 20% - България) за управление на ИТ технологиите.

По отношение на **очакванията за промени в използването на ИТ услуги в университетите**, почти единодушно анкетиранияте посочиха че очакват **тяхното търсене да нараства с времето** (89% - Балкани, 91% - България). Според тях, доминантно те ще са по-скоро умерени към значителни (фиг. 2), като тук отново констатираме различие в очакванията на регионално и национално ниво. Ако като цяло балканските университети очакват един умерен темп на промени, то у нас очакванията са за значително по-радикални такива. Като цяло следва да се отбележи, че в сравнение с аналогично изследване в САЩ (The Chronicle of Higher Education, 2014, с. 18), очакванията в изследвания регион за радикалността на предстоящите технологични промени в университетите са значително по-умерени.



Фиг. 2. Очаквания за предстоящи технологични промени в университетите

Успоредно с това, ИТ мениджърите осъзнават, че университетите са изправени пред редица предизвикателства и промени, с които следва да се справят в новия дигитален век. Сред най-важните **тенденции, които се очакват да въздействат върху дейността на университетите**, те открояват **образование, базирано на компетенциите** (59% - Балкани, 45% - България), **успеха на студентите** (48% - Балкани, 45% - България), **електронните изследвания** (45% - Балкани, 36% - България) и **ролята на институционалния бранд** (41% - Балкани, 45% - България). Значителен приоритет, особено в България, анкетиранияте дават и на **преосмисляне на бизнес модела** (36%), както и на **иновативните образователни пространства** (36%).

Що се отнася до **ефекта от навлизането на новите дигитални технологии в университетите**, анкетиранияте свързват своите очаквания с **повишаване на реализираната печалба чрез оптимизиране и подобряване на операциите** (17% - Балкани, 27% - България), **намаляване на разходите** (17% - Балкани, 9% - България), както и **увеличаване на бизнес възможностите** чрез механизмите на дигиталните канали (14% - Балкани, 18% - България), като университетите у нас посочват и значението на новите технологии за **създаване на тесни партньорски отношения** (18%).



Фиг. 3. Образователни иновации с позитивно въздействие върху дейността на университетите

Важен аспект от дейността на съвременните университети е въвеждането на образователни иновации, с оглед повишаване на качеството на обучение, удовлетвореността на студентите и конкурентните предимства на институцията. Запитани „Кой според Вас ще са иновациите, които ще имат най-позитивно въздействие върху висшето образование в бъдеще“, анкетираните университети посочват **адаптивното обучение за персонализирано образование (28%), хибридните курсове (24%) и отворените образователни ресурси (17%)**. У нас картината е малко по-различна, като университетите акцентират своите очаквания приоритетно към **образование базирано на компетенции (27%), адаптивно персонализирано обучение (18%) и безплатни и отворени образователни ресурси (18%)**.

По отношение на **текущото използване на информационните технологии в образователния процес**, анкетираните посочват, че те се използват за **подобряване процеса на усвояване на знания от студентите (54% - Балкани, 44% - България)** или тепърва протича процес по преосмисляне на тяхното използване в образователния процес (31% - Балкани, 33% - България). Освен това, **университетите в региона определено изразяват предпочитания пред хибридната форма на обучение (87% - Балкани, 75% - България)** пред изцяло неприсъственото онлайн обучение, а що се отнася до налагащите се **МООС (Massive open online courses)**, те не са обект на разглеждане и внедряване в преобладаващата част от изследваните университети (50% - Балкани, 43% - България). Решенията, свързани с **управлението на образователните иновации**, включително и чрез използването на информационни технологии, е **концентрирано основно в специализирано звено за онлайн обучение (42% - Балкани, 56% - България)** и в много по-редки случаи се взема на по-високи управленски нива.

В чисто **технологичен план**, очакванията за настъпващи промени в дейността на университетите в краткосрочен план са свързани предимно с **навлизането на мобилните технологии и социалните мрежи (59%), модернизация на бизнеса (41%), интернет на нещата (41%), управлението и съхраняването на данни (38%) и облачните услуги (38%)**. В този аспект на изследването, очакванията на **българските ИТ мениджъри** се различават, като те свързват своите професионални предизвикателства с области като **мобилните технологии и социалните мрежи (73%), клауд компютинга (55%) и съхраняването и управлението на данни (55%)**, както и процесите по **оптимизация на сървърната инфраструктура чрез нейната виртуализация (45%)**, и не на последно място, предизвикателствата по **гарантиране на сигурността на информационните системи и данни (45%)**.



Фиг. 4. Технологични иновации с очаквано значимо въздействие върху дейността на университетите в краткосрочен план

Същевременно, анализът показва, че **приоритетното целево ИТ финансиране** в университетите е свързано с процесите по **модернизация на бизнеса** (31% - Балкани, 27% - България), **анализи и бизнес разузнаване** (31% - Балкани, 27% - България), **информационна сигурност** (28% - Балкани, 27% - България) и **управление и съхраняване на данни** (28% - Балкани, 18% - България). С висока степен на приоритетност у нас изпъкват и дейности като оптимизиране на съвърнатата инфраструктура чрез виртуализационни технологии (27%), имплементирането на мобилни технологии и социални мрежи (27%), както и търсене на механизми за прилагане на интернет нещата в образованието (27%).

Що се отнася до налагащите се през последните години **облачни решения** и технологии, изследването констатира известни различия на регионално и национално ниво. Ако като цяло **балканските университети са внедрили** (36%), **работят** (9%) или поне декларират че **облачните технологии попадат в стратегическите им приоритети** (18%), то в **България преобладаващата част не разглежда** (38%) или **проучва** (13%) приложимостта им в университетите, като едва 25% посочват че вече имат внедрени подобни решения.

Независимо от посочените констатации, почти две трети от анкетираните посочват че този тип технологии са приложими в образованието. **Приоритетни области** в тази насока са **миграцията на електронната поща и съхранението на данни и социалните мрежи на университетите**, както и управлението на онлайн курсове и студентски приложения. Сред приоритетите на **българските университети** можем да открием и желанието за **преминаване към облака на съществуващите дейта центрове и библиотечни приложения**.

Що се отнася за **възможността за изнасяне (аутсорсинг) на ИТ дейности**, то **преобладаващата част от анкетиранияте са посочили, че не разглеждат подобна опция** (19% - Балкани, 29% - България). Сред предпочитаните дейности за аутсорсване се открояват разработката на приложения, част от ИТ дейностите, в които университетите нямат достатъчно персонал и експертиза.

В заключение можем да посочим, че въпреки че университетите са едни от създателите на съвременните информационни и комуникационни технологии, тяхното широко навлизане в различни аспекти от дейността им изостава значително спрямо корпоративния сектор. Сред основните причини за това можем да посочим както разнопосочния характер на осъществяваните дейности, приоритетността на академичните над административните дейности, ниската степен на централизация на управлението на институцията (в резултат на академичната автономия), както и липсата на пряк и видим финансов стимул от внедряването на информационните технологии (за разлика от корпоративния сектор). Сред ключови предизвикателства пред навлизането на информационните технологии в различните аспекти на дейността на съвременните университети можем да открием както нуждата от „интеграция на културите“ чрез която да се интегрират различните аспекти от университетската култура в единна информационна среда, позволяваща университета да получи нови конкурентни и стратегически предимства от интегрираното академично управление, така и „добавяне на стратегическа стойност“ за университета чрез редизайн на образователните, научните и бизнес процесите в институцията, с цел адаптиране към предизвикателствата на новата икономика на знанията и променящите се образователни и бизнес модели на функциониране на висшите училища.

References

1. Донева, Р., & Касъклиев, Н. (2006). *Стандартизирана университетска информационна система*. V-та Национална конференция „Образованието в информационното общество”.
2. Казакова, И.А. (2011). *История вычислительной техники*. Пенза: Издательство ПГУ.
3. Лазаров, Л. (2016). *Информационни системи в образованието*. Велико Търново: Университетско издателство „Св. Св. Кирил и Методи”.
4. Caspary, K., Kusserow, T., Lavin, J., & Movassaghi, M. (1999). *Managing Technology Efficiently in California K-12 Schools: Policies & Practices for Minimizing the Total Cost of Ownership (TCO)*. Berkeley: University of California.
5. Nicolet, T. (2011). *Leadership in Higher Education: The CIO Role and the Leadership Team*. Dissertation. Greensboro: University of North Carolina.
6. Penrod, J., Dolence, M., & Douglas, J. (1990). *The Chief Information Officer in Higher Education*. Professional paper Series N4, CAUSE.
7. Quinn, M. J. (2015). *Ethics for the information age*. New Jersey: Pearson Education, Inc.
8. Sanyal, B.C. (1995). *The use of computerized information systems to increase efficiency in university management*. IIEP Contributions No. 20, International Institute for Educational Planning, Paris: UNESCO.
9. Selingo, J. (2014). *College 2.0. How Chief Information Officers and Faculty View Technology and the Future of Higher Education*. The Chronicle of Higher Education.
10. Suess, J. (2016). *The Transforming Role of the Higher Ed Chief Information Officer. The Evollution – A Destiny Solutions Illumination*. Изтеглено от http://evollution.com/managing-institution/operations_efficiency/the-transforming-role-of-the-higher-ed-chief-information-officer/.

11. The Chronicle of Higher Education. (2014). *College 2.0: How Chief Information Officer and Faculty View Technology and the Future of Higher Education*. Washington: The Chronicle Of Higher Education.
12. Ward, J., & Peppard, J. (2002). *Strategic Planning for Information Systems*. West Sussex: John Wiley & Sons Ltd.
13. Wikipedia. (2014). DADRA. Изтеглено от <https://bg.wikipedia.org/wiki/DARPA>

VOICE RECOGNITION - HISTORICAL DEVELOPMENT AND MAIN TECHNIQUES

Hasanov Hasan, Burgas Free University, hasan.mehmedov@gmail.com

Georgieva Penka V., Burgas Free University, pgeorg@bfu.bg

Abstract: The natural language processing is one of the main areas of modern artificial intelligence. Voice recognition is an element of natural language processing and aims at transforming spoken words into written text by various techniques. Researchers in area of voice recognition face many challenges that have different sources.

In this study the historical development of voice recognition is presented, the voice recognition types are and the basic techniques used in the field are outlined.

Keywords: voice recognition, speech recognition, natural language processing, artificial neural networks, artificial intelligence

ГЛАСОВО РАЗПОЗНАВАНЕ - ИСТОРИЧЕСКО РАЗВИТИЕ И ОСНОВНИ ТЕХНИКИ

Хасан Хасанов, Бургаски свободен университет, hasan.mehmedov@gmail.com

Пенка В. Георгиева, Бургаски свободен университет, pgeorg@bfu.bg

Абстракт: Обработването на естествени езици е една от основните области на съвременния изкуствен интелект. Гласовото разпознаване е елемент на обработката на естествени езици, при който изречените думи се преобразуват в писмен текст с помощта на различни техники. Разпознаването на глас е област, в която изследователите се изправят пред множество предизвикателства от разнообразен характер.

В тази студия е направен обзор на историческото развитие на гласовото разпознаване, посочени са видовете гласово разпознаване и са представени основните техники, използвани в тази област.

Ключови думи: гласово разпознаване, обработка на естествени езици, изкуствен интелект

I. ВЪВЕДЕНИЕ

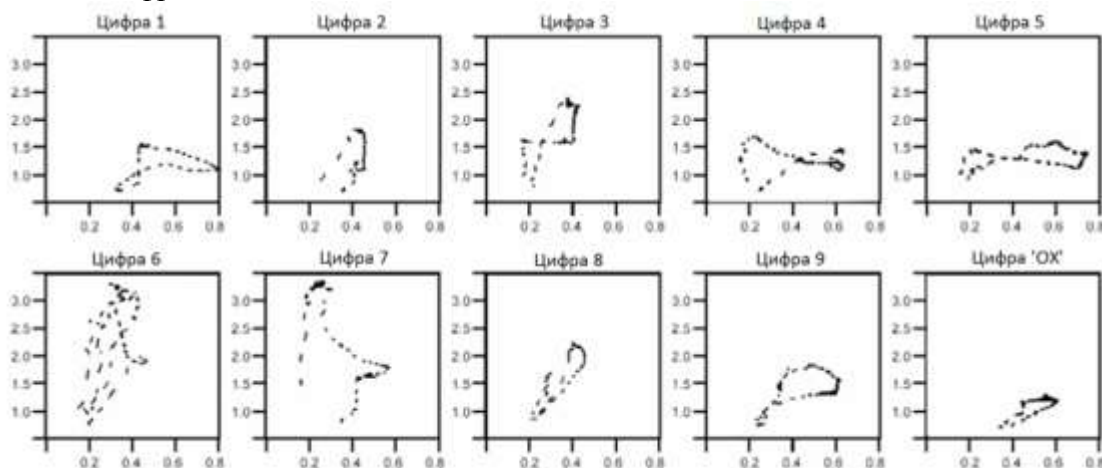
Понятието *гласово разпознаване* (още познато като *разпознаване на глас*, *автоматично гласово разпознаване* и *компютърно разпознаване на глас*) се използва основно в два аспекта: 1) превръщането на изречените думи в писмен текст и 2) обучаването на дадена софтуерна система да разпознава конкретен глас.

Приложенията за разпознаване на глас включват гласови потребителски интерфейси, като гласово набиране (пр. *Обади се вкъщи!*), обработка на обажданията (пр. *Бих искал да проведа разговор.*), управление на домашните електроуреди, търсене (пр. намиране на подкаст, когато определени думи са изговорени), просто въвеждане на данни (пр. въвеждане на номер на кредитна карта), подготовка на структурирани документи (пр. доклади), гласова реч при обработка на текст (пр. текстообработка или електронни писма) и други.

II. ИСТОРИЧЕСКО РАЗВИТИЕ

Първи автоматични системи за речево разпознаване

Ранните опити за създаване на системи с автоматично речево разпознаване са в периода на 60-те и 70-те години на XX век. Те са силно повлияни от теорията за акустичната фонетика, която описва фонетичните елементи на речта (основните звуци на езика) и обяснява как тези елементи са акустично реализирани при говоримото изказване. Основните звуци са фонемите, но съществена роля имат и съответното им място и начинът на артикулация, използван за възпроизвеждане на звуците в различни фонетични контексти. Например, за пресъздаването на стабилен звук на гласна, гласните струни трябва да вибрират, а въздухът, който се разпространява през вокалния тракт се възпроизвежда като звук с естествени резонансни режими, аналогични на тези, които се появяват при акустичната туба. Тези естествени резонансни състояния се наричат форманти (формантни честоти). На фигура 1 са изобразени формантните траектории на първа и втора формантна честота при произнасяне на всяко от цифрите. Тези траектории служат за референтен модел при определянето на идентичността на изговорена непозната цифра. [1].



Фиг. 1. Сравнение на възпроизвеждането на формант 1 и формант 2 на цифрите 1, 2, ..., 9 и непозната цифра OX

При други ранни системи за разпознаване, Олсън и Белар от лабораториите RCA създават система за разпознаване на 10 срички от един конкретен говорещ [2], а в Линкълн Лаб на Масачузетския технологичен институт, Дж. Форги и К. Форги проектират независима от говорителя система за 10-гласно разпознаване [3]. По същото време, независимо от тях японски изследователи създават специализиран хардуер, който да изпълнява задача по разпознаване на реч. Най-важни са разработките на Сузуки и Наката в Изследователската радио лаборатория в Токио [4], на Сакаи и Дошита от университета в Киото [5] и цифровия разпознавател на лабораториите NEC [6]. Работата на Сакаи и Дошита включва първата употреба на речеви сегментатор за анализ и разпознаване на реч в различни дялове от подадения изказ. Разработката на университета в Киото може да се счита като предшественик на системата с непрекъснато разпознаване на гласа.

Друга система за разпознаване създава Дийнс (Университетски колеж, Англия) и това е фонемен разпознавател, който разпознава 4 гласни и 9 съгласни [7]. Включвайки статистически данни за валидните фонемни последователности в английския език, те увеличават цялостната фонемна разпознаваща точност за думите, съдържащи две или

повече фонемни. В тази разработка за първи път е употребен статистически синтаксис на фонемно ниво при автоматично речево разпознаване.

Алтернатива на употребата на речеви сегментатор е концепцията за въвеждане на разнородна времева скала за изравняване на речевите модели, предложена независимо от Т. Мартин от лабораториите RCA [8] и Винчук от Съветския съюз [9]. Мартин открива необходимостта за справяне с времевата разнородност в повтарящи се речеви сегменти и предлага набор от решения, включително маркиране на крайните точки на изговорената последователност, което в голяма степен подобрява надеждността на действието на разпознавателя. Винчук предлага използването на динамично програмиране за времето между две изказвания, за да извлече състоятелна оценка за тяхната прилика. Неговата разработка, макар и в голяма степен непозната на запад, предшества тази на Сакое, Чиба [10] и други, представили по-формални методи, познати като динамични времеви изкривявания, при съгласуване на речевите шаблони. След публикацията на Сакое и Чиба, динамичното програмиране, в множество разнообразни форми (включително алгоритъма на Витерби [11]), се превръща в задължителна техника при автоматичното речево разпознаване.

Технологично развитие на комерсиални системи

Атал и Итакура независимо един от друг представят основните концепции на LPC (Linear Predictive Coding) [12], [13], като процедура за ефективно кодиране на речевата вълна, която е представена чрез времеви параметри, свързани с реакцията на вокалния тракт. До средата на 1970-те, основните идеи на тази технология са разработвани от Итакура [14], Рабинер и Левинсън [15] и др.

По същото време Том Мартин основава първата комерсиална компания Threshold Technology Inc., предлагаща софтуерна система за речево разпознаване - VIP-100 System. Системата се използва единствено в няколко приложения от фирми за производство на телевизионни фасадни плочи за контрол на количество и от FedEx за сортиране на пакети на конвейерна лента, но главната ѝ значимост е начинът, по който указва влияние върху Агенцията за напреднали изследователски проекти (Advanced Research Projects Agency, ARPA) да започне финансиране по програмата за изследването на речево разпознаване Speech Understanding Research (SUR). Сред системите, създадени по тази програма е *Hapry* на университета Карнеги Мелон [16], която доказва способността си да разпознава реч, използвайки речник от 1011 думи, при това със сравнително голяма точност. В системата *Hapry*, входната реч, след като премине през параметричен анализ, се сегментира, след което сегментираната параметрична последователност се подлага на сравнение с телефонен шаблон, използвайки разстоянието на Итакура. Търсенето по графи компилира, предполага, изравнява и след това потвърждава разпознатата последователност от думи (или звуци), която удовлетворява ограниченията с най-високия възможен резултат. *Hapry* е първата система, която се използва мрежа от крайни състояния, за да намали изчисленията и ефективно да определи най-близкия съвпадащ низ. Методи за оптимизиране на получената крайна мрежа от състояния се появяват едва след 1990 г. [17]

Други системи, разработени под програмата SUR на ARPA са *Hearsay-II* на CMU и *HWIM* на BBN [18]. Нито една от двете системи не успява да реализира амбициозната цел на ARPA за ефективност. Въпреки това, подходът, предложен в *Hearsay-II* за използване на паралелни асинхронни процеси, симулиращи компоненти от източниците на знание в речевата система, е една пионерска концепция. В системата *Hearsay-II* знанията от паралелни източници се интегрира за да се получи следващото ниво на хипотеза. Системата *HWIM* се характеризира с мрежа за лексикално декодиране, обхващаща сложни фонетични правила с цел достигане на точност при разпознаване на фонемите,

способност за справяне с двусмислеността чрез решетка от алтернативни хипотези и концепция за верификация на думи на параметрично ниво. Друга комерсиална система е *DRAGON* на Джим Бекър, който стартира компания със същото име.

Паралелно с проектите на ARPA се оформят две обширни направления в разпознаването на реч и това са разработките на IBM и AT&T Bell Laboratories. Усилията на IBM, водени от Фред Джелинък, са насочени към създаването на „гласово активираща се пишеща машина“ (voice-activated typewriter, VAT), чиято основна функция се свежда до преобразуване на изговорено изречение в последователност от букви и думи, които да бъдат показани на екран или отпечатани на хартия [19]. Създадената разпознавателна система *Tangora* е силно зависима от говорителя - пишещата машина трябва да бъде тренирана от всеки индивидуален потребител. Технологичният фокус е върху познавателната лексика с основен акцент върху офис кореспонденцията и структурата на езиковия модел (граматиката), която е представена от статистически синтактични правила, описващи вероятността за появяване в речевия сигнал на последователността от езикови символи (т.е. фонемите или думи). Прието е такъв тип речево разпознаване да се нарича *транскрипция*. Наборът от статистически граматически или синтактични правила се нарича *езиков модел*, като най-често използван е *n-gram* моделът. Въпреки че *n-gram* езиковият модел и традиционната граматика са проявления на езиковите правила, техните роли са фундаментално различни. Моделът *n-gram* характеризира връзката между думите в обхвата от *n* на брой думи и в този случай статистическо представяне на граматиката е удобно средство. Ефикасността му при търсенето по думи е демонстрирана в световно известната игра на Клод Шанон [20], в която човекът и компютърът трябва да познаят каква ще е следващата дума в произволно изречение. Човекът познава на базата на вродения опит с езика, а компютърът използва натрупаните статистики с думи, за да направи своя най-добър избор, базиран на максималната вероятност от изчислените честоти за поява на дадена дума. Демонстрирано е, че веднъж, след като наборът от думи надвиши 3 ($n > 3$), много по-вероятно е компютърът да победи човека. След тяхното въвеждане през 80-те, употребата на *n-gram* езиковите модели и на техните варианти, се превръща в неизменна част от системите за разпознаване на реч.

От друга страна, изследователската програма на AT&T Bell Laboratories има за цел да се предоставят автоматизирани телекомуникационни услуги на обществото (пр. гласово набиране) и управление на маршрутизирането на телефонните обаждания. Очакването е тези автоматични системи да работят добре за десетки милиони потребители без нуждата от индивидуален тренинг за всеки говорещ. Основният акцент на Bell Laboratories е системата да е независима от говорителя и да може да се справи с акустичната променливост, присъща на речевите сигнали, идващи от различни говорещи, често с коренно различни местни акценти. В следствие изследването на акустичното разнообразие на различни говорители води до изучаването на съвкупност от измервания на спектралните разстояния и техники за статистическо моделиране [21], които в резултат дават значително точна репрезентация на произношенията на голям брой хора. Основна използвана техника при непрекъснатото разпознаването на речта е скрития модел на Марков [22], [23]. Тъй като приложения като гласово набиране и маршрутизиране на разговори обикновено изискват късо изговаряне на ограничена лексика и се състоят само от няколко думи, акцентът в изследването на Bell Laboratories се поставя по-скоро върху това, какво всъщност се има предвид под *акустичен модел* (т.е. спектралното възпроизвеждане на думи или звуци) отколкото върху същността на езиковия модел (т.е. представянето на граматиката или синтаксиса). Освен това, от голямо значение в подхода на Bell Laboratories е концепцията за *търсене на ключови думи* като примитивна форма за разбирането на речта [24]. Техниката за търсене на

ключови думи изисква разширяване на типичната парадигма за разпознаване на шаблони с такава, която включва тестване на хипотези.

Подходите на IBM и AT&T Bell Laboratories имат обща характеристика – и при двата ключова роля има математическият формализъм.

Системи за гласово разпознаване, използващи техники на изкуствения интелект

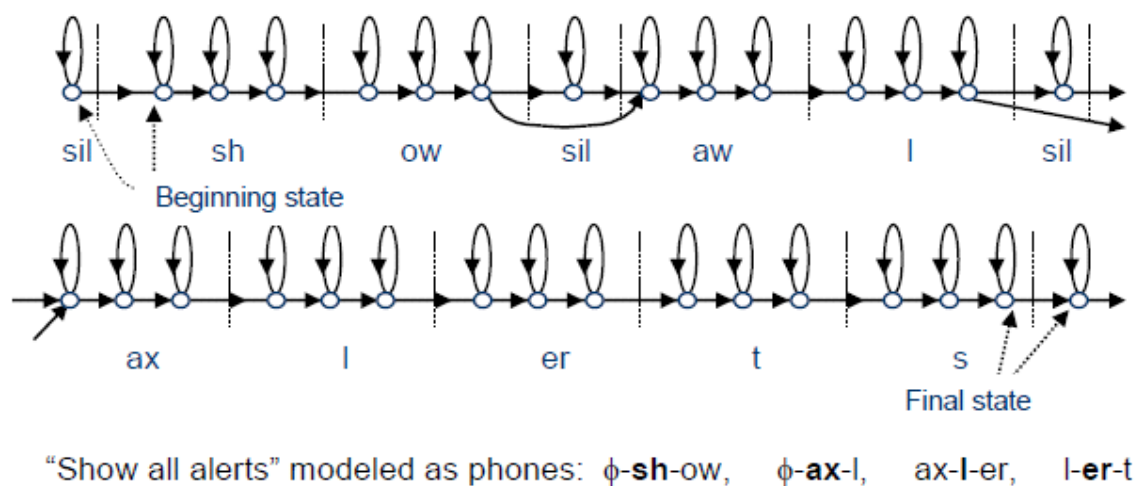
След 1980 г в областта на речевото разпознаване се наблюдава промяна на методологията от интуитивно-шаблонно ориентиран подход към статистическо моделиране, базирано на скрит модел на Марков (*СММ*). Основната идея на *СММ* е бече използвана от лабораториите на IBM и IDA [25] [26], но методологията е доразвита и се превръща в предпочитан метод за речево разпознаване, като тази тенденция остава постоянна през следващите две десетилетия.

СММ оперира със стохастични процеси и са подходящи за моделиране както на присъщата променливост на речевия сигнал и получаващите се спектрални особености, така и на структурата на говоримия език [27]. Реалният речеви сигнал по своята същност е изключително променлив поради вариации в произношението и акцента, както и заради фактора околна среда, като ехо и шум. Когато хората изговарят една и съща дума, акустичните сигнали не са идентични (всъщност могат да бъдат удивително различни), макар основната лингвистична структура по отношение на произношение, синтаксис и граматика може да останва същата. Формализмът на *СММ* използва верига на Марков за представяне на лингвистичната структура и лингвистични разпределения за променливостта на акустичното представяне на звуците. При даден набор от познати изказвания, представляващи достатъчна съвкупност от вариации на желаните думи (наречени тренировъчен набор), може да се използва ефикасен оценъчен метод, наречен алгоритъм на Баум-Велх [28], за да се получи „най-добрия“ набор от параметри, определящи съответстващия модел. Оценката на параметрите, дефиниращи модела е еквивалентна на трениране и обучение. Вероятностната мярка, представена от скрития модел на Марков е важен компонент при системите за речевото разпознаване, които следват подхода за разпознаване на статистически модели, и води началото си от теорията на Бейс за взимане на решение [29].

Идеята за използване на *СММ* възниква в Института за анализ на отбраната в Принстън. Баум определя *СММ* като набор от вероятностни функции на верига на Марков, която по определение включва две вложени разпределения, като едното се отнася до верига на Марков, а другото – до набора от вероятностни разпределения, всяко от които е съотнесено към състояние от верига на Марков [30]. Този двойно стохастичен процес е полезен в приложения за предвиждане на фондовия пазар и крипто-анализът на ротационен шифър, използван по времето на Втората световна война. Техниката на Баум за моделиране и оценка първо се прилага върху дискретни наблюдения, а след това случайните наблюдения се моделирани функции на вероятностната плътност. Двойно стохастичният процес на Баум намира приложение в системите за разпознаване на говор [31] [32], но по-късно става ясно, че ограничението на вида на функциите на плътността налага намаляване на производителността на системите, особено при задачи, независими от говорителя. В лабораториите Bell Laboratories въвеждат *СММ* със смесени вероятностни функции на плътността [33] [34], които оттогава нататък се утвърждават като изключително важна част от осигуряването на задоволителна точност при разпознаването, особено при лексикално големи задачи за разпознаване на реч, независими от говорителя.

СММ е неизменна част от големи декодиращи речта структури с включен езиков модел. Употребата на граматика от крайни състояния при продължителното речево разпознаване на голяма по обем лексика представлява разширяване на *СММ*, при което

структурата на езика се отчита на ниво взаимодействие между артикулацията и произношението. Въпреки, че тези структури са груби приближения на реалния речеви процес, изчислително те са ефективни и често достатъчни, за да се получи смислен начален резултат на изпълнение. Сливането на *СММ* (със своето предимство в статистическата съгласуваност, в частност при работа с акустичното разнообразие) и мрежата на крайни състояния (с нейното търсене и изчислителна точност, по-точно при работа с хипотези за последователност от думи) е важно технологично развитие през средата на 80-те години на XX век.



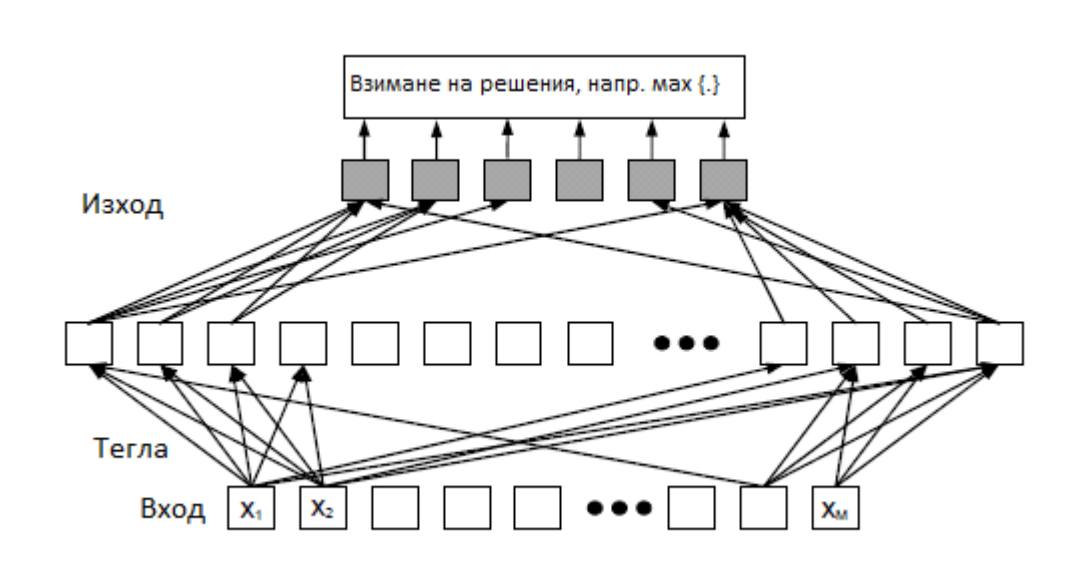
Фиг. 2. Комбинирана мрежа на крайните състояния за изричането на “show all alerts”

На фигура 2 е показан комбиниран модел на крайните състояния за изказването на “show all alerts”, съставен от няколко модели на части от думата, които зависят от контекста и представляват съответстващите фонемо-подобни единици (включително единица за тишина *sil*, която може да се появи в началото или в края на изречението, както и в края на всяка дума в изречението, подобно при пауза по време на говорене). Графът на крайните състояния се реализира като верига на Марков за пресмятане на вероятността от спектрално представяния във времето наблюдения на неизвестна реч. Всеки възел в графа се асоциира с вероятностно разпределение, което отговаря на променливостта при реализиране на съответния фонемо-подобен звук. Вероятността изказаното да е генерирано от мрежа на крайните състояния, представена от модела, се изчислява като последователна сума от локални вероятности, отнасящи се до елементарните единици на съставния модел, след като се изпълнява изравняване на състоянията чрез динамично програмиране, за да се максимизира попадението между означените единици и съответстващите части от наблюденията на речта дори и за модели с неправилни последователности от думи.

Във всеки интервал от време има набор от хипотетични единици, а определянето на еднаквостта на звука се основава върху възможно най-високата вероятностна стойност или резултат на съвпадение. Броят хипотетични единици за съвпадение и този на пътищата за търсене може да бъде много голям и да изисква ефикасни изчислителни алгоритми, за да се намери решение на задачата. В средата на 90-те години се разработва средство, наречено библиотека на крайните състояния (*FSM=finite-state machine library*,

което се превръща в неизменна част от всички модерни системи за разбиране на текст и речево разпознаване [35].

Друга технология, използвана обработката на естествени езици е идеята за изкуствена невронна мрежа. Невронните мрежи са представени още през 1943 г., но в началото не показват задоволителни резултати [36]. След появата на модела за паралелна разпределена обработка (parallel distributed processing), и по-конкретно на многослойния персептрон (фиг. 3), използването на невронни мрежи придобива широка популярност. Причината не е в аналогията с дейността на нервната система, а поради способността за апроксимация на произволна функция, при условие, че няма зададено ограничение за сложността на обработващата конфигурация. Ако разпознавателят на шаблони се онагледява като средство, извършващо свързване на функцията от входния модел с идентификатор от класа, тогава може да се каже, че многослойният персептрон би бил идеално средство за тази задача. Основната цел на първоначалните невронни мрежи за разпознаване на речт е съсредоточена върху прости задания като разпознаване на няколко фонеми или малък брой думи. Тъй като разпознаването на говора неизменно изисква боравене с изменения във времето, в своя ранен стадий невронните мрежи не са достатъчно добро средство [37]. Научноизследователска дейност по това време има за цел да интегрира невронните мрежи с главната структура на скрития модел на Марков.



Фиг. 3. Многослоен персептрон

През 90-те години задачата за разпознаването на шаблони постепенно се преобразува в оптимизационна задача за минимизиране на емпиричната грешка [38], защото главната цел при проектирането на разпознавателя е достигането на минимална грешка при разпознаване, а не най-добро приспособяване на функцията на разпределение към данните според критерия на Бейс. Концепцията за минимална емпирична грешка последователно дава като резултат набор от техники, една сред които е дискриминиращото обучение и базираните на ядра методи, като помощно векторните механизми (*SVM= support vector machines*) [38], [39].

Успехът на статистическите методи съживява интереса на ARPA към системите за разпознаване, включително системата *Sphinx* на CMU [40], системата *BYBLOS* на BBN [41] и системата *DECIPHER* на SRI [42]. В системата *Sphinx* успешно е интегриран

статистическия метод на *СММ* със предимствата на по-ранната система *Harpy* за търсене в мрежа. В резултат системата обучава и въвежда контекстово зависими телефонни модели в сложна лексикална декодираща мрежа, постигайки значителна точност при речевото разпознаване на голям брой от думи.

В ARPA полагат усилия и за създаване на критерии за оценка на технологиите за разпознаване на говор. Подобни оценки главно използват мерки за степента на грешка в дума и изречение, както и обхвата на системата за разпознаване. Оценки систематично се прилагат върху внимателно проектирани задания с нарастващо ниво на сложност, вариращи от разпознаване на последователен говор, произнесен със стилизирана граматическа структура (пр. при управление на военноморски ресурси) до транскрипции на живи телевизионни предавания и на ежедневна реч (пр. NAB включва лексика от над 20 млн. думи). Една от разработките е *ATIS* - задача, включваща прост спонтанен речеви разговор с автоматизирана система за извличане на информация от въздушния транспорт. Друга разработка *WSJ* се отнася до транскрипцията на набор прочетени параграфи от списанието *Wall Street*, като обемът на лексиката достига до 60 млн. думи. Най-голямото предизвикателство, предложено от ARPA е заданието *Комуникационно табло (Switchboard)*. Речта е разговорна и спонтанна, с много инстанции на т. нар. несигурности, като например непълни думи, колебание, поправяне и т.н. Основното заключение, което може да се извлече от тези резултати е, че получаването на очаквания резултат при обработване на разговорна реч, която не е задължително да се подчинява на лингвистични ограничения, с значително по-трудно от обработване на реч, която следва стриктни граматически и семантични правила. В допълнение, оценъчната програма показва, че увеличението на количеството речеви данни, използвани за изчисление на разпознавателните параметри (т.е. размерът на обучаващата серия) винаги води до намаляване нивото на грешката в думите, като е прието нивото на грешка да бъде под 10%.

След 90-те години се наблюдава съществен подем на много изследователски програми по света. Всички системи за обработка на естествени езици стават по-сложни и точни. Добре структурираната базова софтуерна система е незаменима в по-нататъшното проучване и развитие за комбинирането на нови концепции и алгоритми. Системата *Hidden Markov Model Tool Kit (HTK)*, разработена от екипа на университета Кеймбридж с ръководител Стийв Янг, е една от най-широко използваните софтуерни системи при проучванията за автоматичното разпознаване на речта [43].

Комуникация човек-машина

В ранните етапи от развитието на приложенията за разпознаване на реч изследователите се фокусират върху преобразуване на акустичната реализация на лингвистично събитие в думи. Концепцията, че процесът реч-към-текст е задължителна първа стъпка от комуникацията човек-машина, има своите последователи и до днес. Но при такава комуникация потребителите обикновено изговарят изречения, които в повечето случаи не удовлетворяват напълно граматическите ограничения на разпознавателя и включват нелексикални, а изговорените изкази често са влошени от лингвистично ненужни шумови компоненти като заобикалящ шум, неприсъщи акустични звуци, смущаваща реч и др. Освен това, както и при комуникацията човек-човек, употребата на речта често изисква диалог между потребителя и машината, за да се постигне някакво удовлетворително ниво на разбиране. Подобен диалог често изисква операции като заявка и потвърждение. Методът за търсене на ключови думи е решение за първия проблем, докато вторият насочва вниманието на изследователската общност върху обработката на диалога. Приложения, в реалистично се пресъздават комуникационните

възможности на човека са например системите *Pegasus* и *Jupiter* на екип от Института по технологии в Масачузетс с ръководител на Виктор Зю [44] [45] и системата *How May I Help You (HMHU)* на AT&T, дело на Ал Горин, която от 2000 година е неизменна услуга при обслужването на клиенти на фирмата [46].

Pegasus е система за разговорна реч, която осигурява информация за състоянието на въздушната линия над обикновена телефонна линия. *Jupiter* е подобна система, фокусирана върху достъпа до информация за локално и национално метеорологичното време. С добре проектирано управление на диалога, тези системи насочват потребителя да предостави нужната за изпълнение на заявката, сред малък и безрезервен набор от меню с опции, без излишно да искат детайли за заявката. Управлението на диалога често включва вградена информация за разпознати фрази и софт поправяне на грешки, така че потребителя да реагира така, сякаш в другия край на телефонната линия има реален човешки посредник, а не машина. Целта е да се проектира машина, която комуникира, а не само разпознава думи от изречен изказ.

Въпреки, че автоматичното разпознаване на говор и системите за разбиране на реч не са напълно свършени по отношение на прецизността на думите/заданията, добре развитите приложения се възползват от съществуващата технология, за да предоставят реална стойност на клиента и броят и обхвата на такива системи постоянно се увеличава.

III. ВИДОВЕ ГЛАСОВО РАЗПОЗНАВАНЕ

Дискретно разпознаване на глас

При дискретното разпознаване на глас (*Discrete Voice Recognition*) се използва основно техниката *разпознаване на изолирани думи (Isolated Word Voice Recognition)*. Разпознаването на изолирани думи е процес, при който след всяка дума трябва да бъде направена пауза за отделяне. Паузата, или липсата на звук е основен подход за определяне на началото и края на изговорена дума.

Дискретното разпознаване на глас извършва преобразуване на глас в текст дума по дума. В по-напредналите версии на дискретно гласово разпознаване се конвертират цели фрази, като това обикновено е едно изречение. Дискретно разпознаване на глас се използва предимно в приложения за диктовка и при гласова навигация с команди.

Непрекъснато разпознаване на глас

При непрекъснатото разпознаване на глас (*Continuous Voice Recognition/Connected Words Voice Recognitions*) началото и края на думата може да бъде определено, без да е необходима звукова пауза е вид разпознаване. Този вид разпознаване е сложна форма на разпознаване и изисква много по-голям ресурс.

Непрекъснатото разпознаване на глас също се нарича свързано гласово разпознаване, тъй като не изисква значителна пауза между думите. Вместо това говорителят може да говори по-естествено и въпреки това софтуерът за гласово разпознаване разбира къде започва думата и къде свършва.

Спонтанно гласово разпознаване на реч

Спонтанното гласово разпознаване на речта (*Spontaneous Speech Voice Recognition/Natural Speech Voice Recognition*) е осъвършенствана форма на непрекъснатото разпознаване. Несъзнателни звуци, които хората издават, докато говорят като едновременно мислят и оформят нерепетирани изречения, могат да бъдат разбрани от софтуера и съответно да бъдат пропуснати при преобразуването. Ползата от спонтанното

гласово разпознаване на речта е, че не е нужно да се променя начина на говорене с компютъра.

Спонтанното гласово разпознаване на говор се нарича още "естествено разпознаване на гласа", тъй като поддържа разпознаване на естествения начин, по който говори човек.

Разпознаване на глас, зависещ от говорещия

Разпознаване на глас, зависещ от говорещия (*Speaker Dependent Voice Recognition*), е вид разпознаване, който зависи от гласа на говорещия човек. Този вид разпознаване изисква обучение, за да стане преобразуването на речта в текст по-точно. Обучението се извършва от говорителя. Резултатът е система за гласово разпознаване, която разбира специфичния акцент и глас на съответния човек.

Системите за разпознаване на глас, които зависят от говорителя, могат да бъдат дискретни или непрекъснати.

Разпознаване на глас, независещ от говорещия

Разпознаване на глас, независещ от говорещия (*Speaker Independent Voice Recognition*) е тип разпознаване на глас, което е обратното на разпознаване на глас, зависещ от говорещия. Този вид разпознаване не изисква обучение и може да разбере реч от широк кръг говорители.

Системите за разпознаване на глас, независими от говорещия, изискват от системите повече възможности за обработка и в общия случай не са толкова точни при преобразуването на речта в текст.

Разпознаване на глас от естествен език

Разпознаване на глас от естествен език (*Natural Language Voice Recognition*) всъщност е добавка към непрекъснатото спонтанно разпознаване на глас и се отнася до способността на компютъра да разбере въпрос или команда, казана по естествен начин, а не по структурирания начин, който очаква системата. Обикновено това се отнася до способността на компютъра и да отговаря на говорещия с естествен език, като че ли се провежда разговор.

Естественото разпознаване на глас е това, което управлява съвременните виртуални асистенти като *Сири*, *Алекса* или *Кортана*.

IV. ТЕХНИКИ ЗА ГЛАСОВО РАЗПОЗНАВАНЕ

Смисловата обработка на естествения език (която включва анализ и синтез) първоначално се свързва със ситаксиса и синтактичния анализ. Скоро обаче се установява, че оставането само на синтактично равнище не дава задоволителни резултати, защото понякога прилагането на едно или друго синтактично правило зависи от смисъла на самия текст. Появява се необходимостта от синтактично-смислов анализ и разбиране на текста и следователно извличане на знанията от текста и представянето им в някаква система. При решаването на този проблем се налагат редица ограничения, например ограничения върху предметната област, на знанията, на допустимите конструкции в езика, на обема на анализираната информация и т.н.

Морфологичен анализ

Морфологичният анализ означава определяне на думата като част на речта и на всички нейни граматически характеристики. Всички форми на думата образуват нейната парадигма. Промяната на думата от една словоформа в друга става чрез промяна на структурата на думата, т.е. чрез прибавяне и отнемане на морфемите (суфикси). Морфемата е най-малката значеща единица в езика.

Думата морфема има гръцки произход и означава значимата част на една дума (корен, представка, наставка, вставки и окончания). При морфологичния анализ се анализират морфемите на думите, т.е. техните компоненти и признаци, и неучастващите в думите признаци (например препинателните знаци – пунктоация) се отделят от думите [47].

Синтактичен анализ

Синтаксисът е дял от граматиката, който се занимава със строежа на изречението и с неговите части. В тази стъпка на гласовото разпознаване изреченията, които са линейни последователности от думи, се преобразуват в структури, които показват как думите се отнасят една към друга. Тази стъпка на граматически разбор, наречена парсване (*parsing*), превръща линейния списък от думи в изречението в структура, която дефинира представените единици от този списък.

Семантичен анализ

Семантиката е дял от граматиката за значението на думите, на техните части и съчетания. При семантичния анализ на структурите, създадени от синтактичния анализатор, се присвояват стойности. В повечето езикови пространства словосъчетанието “*Безцветната зелена жаба*” ще бъде отхвърлена като семантично неправилна. Тази стъпка съпоставя на отделните думи подходящи обекти от базата знания и трябва да създаде коректни структури, които показват по какъв начин значенията на отделните думи се комбинират едно с друго.

Интегриране в съответен контекст

Значението на дадено изречение може да зависи смислово от изреченията, които го предхождат. Обектите, въведени в изречението, трябва или да са експлицитно определени, или да се отнасят към обекти от предишни изречения. Целият контекст трябва да бъде ясен и това е целта на тази стъпка.

Прагматичен анализ

Структурата, представяща какво е било казано преди, се интерпретира отново, за да се определи, какво всъщност означава.

Обработка на сигнала

Обработката на сигнала представлява взимане на говорима част от езика и да се изрази в последователност от думи. Тук се включва дигитализиране на сигнала и различаване на сегментите на думата, които могат да бъдат асемблирани в цели думи. Езиковите елементи, обработвани като сигнали, се наричат фонемите.

АЛГОРИТМИ

Акустичното моделиране и езиковото моделиране са важни части от съвременното базирано на алгоритми статистическо разпознаване на реч. Скрытите модели на Марков

са широко използвани в много системи. Езиковото моделиране има и много други приложения като интелигентна клавиатура и класификация на документи.

Скрити модели на Марков

Модерна система с общо предназначение за разпознаване на реч се основава на скритите модели на Марков, тъй като речевият сигнал може да се разглежда като отделни фрагменти от стационарния сигнал, т.е. като част от целия сигнал. За малък времеви интервал (пр. 10 милисекунди), говорният сигнал може да се приеме за почти стационарен процес.

От гледна точка на теорията на вероятностите един случаен процес притежава свойство на процес на Марков, ако условното разпределение на вероятностите на бъдещи състояния на наблюдавания процес, при предварително известни текущи и минали състояния, зависи само от състоянието към този момент, и не зависи от миналите такива. С други думи бъдещите състояния на един процес са условно независими от миналите състояния, при развитие на наблюдавания процес. Процес, който притежава това свойство се нарича процес на Марков. Най-популярните процеси на Марков са веригите на Марков. Верига на Марков е процес на Марков, който приема стойности от дискретното множество, наречено пространство на състоянията, като стойността му се променя в точно определени моменти от време. Тези вериги са широко разпространени и използвани в теорията на разпознаване на реч, тъй като заемат много малко памет за моделиране на динамични процеси и освен това са лесни за реализация.

В статистическото моделиране скрит модел на Марков е процес на Марков с неизвестни параметри. Всеки определен параметър е функция на вероятностите на дадено състояние. Всяко състояние може да се възпроизведе, което не позволява да се определи поредицата от състояния до момента на наблюдение. С други думи поредицата от състояния не може да бъде наблюдавана и остава скрита.

В общия модел на Марков, състоянието е директно видимо от наблюдателя и за това вероятните състояния на промяна са единствените параметри. В скрития модел на Марков състоянието не е пряко видимо, но променливите, повлияни от състоянието, са видими.

Скритите Марковски модели са особено известни с приложенията си в разпознаването на реч, ръкопис, жестове, и др.

Динамично програмиране

Разпознаването на реч основано на *DTW (Dynamic Time Warping)* е ефективен метод при разпознаващи системи с малък речник от думи.

Алгоритъм за динамично програмиране [48]

Стъпка 1: Инициализация

$$D(1, 1) = d(1, 1), B(1, 1) = 1, \text{ for } j = 2, \dots, M \text{ и се изчислява } D(1, j) = \infty$$

Стъпка 2: Итерации

при $i = 2, \dots, N$

$$\begin{aligned} &\{ \\ &\quad \text{за } j = 1, \dots, M \text{ се изчислява} \\ &\quad \{ \\ &\quad \quad D(i, j) = \min_{1 \leq p \leq M} [D(i-1, p) + d(p, j)] \\ &\quad \quad B(i, j) = \arg \min_{1 \leq p \leq M} [D(i-1, p) + d(p, j)] \\ &\quad \} \\ &\} \end{aligned}$$

Стъпка 3: Връщане назад и преустановяване на процеса

Оптималното разстояние е $D(N, M)$ и оптималния път е $(s_1, s_2, \dots, s_N)$, където $s_N = M$ и $s_i = B(i + 1, s_{i+1}), i = N - 1, N - 2, \dots, 1$.

Алгоритъм Forward за оценка на скрит модел на Марков [49]

Стъпка 1: Инициализация

$$\alpha_1 = \pi_1 b_1(X_1), 1 \leq i \leq N$$

Стъпка 2: Индукция

$$\alpha_t(j) = \left[\sum_{i=1}^N \alpha_{t-1}(i) a_{ij} \right] b_j(X_t), \quad 2 \leq t \leq T; 1 \leq j \leq N$$

Стъпка 3: Преустановяване на процеса

$$P(X|\Phi) = \sum_{t=1}^N \alpha_T(i)$$

Ако е необходимо да се завърши в крайно състояние,

$$P(X|\Phi) = \alpha_T(s_F)$$

Декодиране на модел на Марков с алгоритъм на Витерби

Стъпка 1: Инициализация

$$V_1(i) = \pi_i b_i(X_1)$$

$$B_1(i) = 0$$

Стъпка 2: Въведение

$$V_t(j) = \max_{1 \leq i \leq N} [V_{t-1}(i) a_{ij}] b_j(X_t), \quad 2 \leq t \leq T; 1 \leq j \leq N$$

$$B_t(j) = \arg \max_{1 \leq i \leq N} [V_{t-1}(i)], \quad 2 \leq t \leq T; 1 \leq j \leq N$$

Стъпка 3: Прекратяване

$$\text{The best score} = \max_{1 \leq i \leq N} [V_t(i)]$$

$$S_T^*(j) = \arg \max_{1 \leq i \leq N} [B_t(i)]$$

Стъпка 4: Връщане назада

$$S_t^* = B_{t+1}(S_{t+1}^*), \quad t = T - 1, T - 2, \dots, 1$$

$$S^* = (S_1^*, S_2^*, \dots, S_T^*) \text{ е най-добрата последователност}$$

Оценяване на параметрите на модел на Марков с Forward-backward алгоритъм на БаумУелч

Стъпка 1: Инициализация

Избира се първоначалната оценка Φ .

Стъпка 2: Е-Стъпка

Изчисляване на спомагателната функция $Q(\Phi, \hat{\Phi})$.

Стъпка 3: М-Стъпка

Изчисляване на $\hat{\Phi}$, така че да максимизира спомагателната Q -функция.

Стъпка 4: Итерация

Полага се $\Phi = \hat{\Phi}$ и алгоритмът се повтаря от стъпка 2 до достигане на желаната сходимост.

Ограничения на скрития модел на Марков

В стандартните скрити модели на Марков има редица ограничения като:

- използване на непрекъснато експоненциално разпределение;
- вероятността на прехода зависи само от източника и предназначението;
- всички наблюдавани състояния са зависими само от състоянието, което ги поражда, без значение от наблюдаваните съседни състояния.

Изследователите са предложили редица техники за справяне с тези ограничения, макар че това не гарантира значителни подобрения в точността на разпознаване на реч, като използване на непрекъснато моделиране, условно независими предпоставки и други.

Изменения на речевия сигнал. Изчисляване на грешката при разпознаване.

Характеристики на речевите сигнали

За постигане на ефективност при разпознаване на реч задължително се взимат под внимание следните три фактора:

- 1) изменения в стила на речевата вълна:
 - a. произнасяне на отделни думи или кратки фрази с пауза между тях;
 - b. непрекъснато произнасяне;
 - c. скорост на говорене;
 - d. сила на речевата вълна - шепнене, нормален говор и викове;
- 2) индивидуални гласови характеристики- всеки глас е уникален, различен и подчинен на физическите особености, които влияят върху дължината, ширината и размера на вокалния тракт, пола, възрастта, диалекта, здравословното състояние, образованието и не на последно място маниера на говорене;
- 3) влияние на околната среда.

Оценка на грешката

При разпознаване на реч основно се разграничават три вида грешки:

- замяна - правилната дума е заменена с неправилна;
- изтриване - дадена дума е пропусната;
- вмъкване - добавена е допълнителна дума.

Задачата за максималното съответствие се състои в съпоставяне на броя на разпознати думи към чроя правилни думи и пресмятане на броя на заменените (*Subs*), изтритите (*Dels*), и вмъкнатите (*Ins*) думи.

Степента на сгрешени думи (*Word Error Rate*) се дефинира като:

$$\text{Word Error Rate} = 100\% * \frac{\text{Subs} + \text{Dels} + \text{Ins}}{N},$$

където N е броят на всички думи в изречението.

Алгоритъм за пресмятане на степента на сгрешени думи (word error rate):

Стъпка 1: Инициализация

$$\begin{aligned} R[0, 0] &= 0 \\ R[i, j] &= \infty \\ \text{if } (i < 0) \text{ or } (j < 0) & B[0, 0] = 0 \end{aligned}$$

Стъпка 2: Итерации

$$\begin{aligned} &\text{for } i = 1, \dots, n \{ \\ &\quad \text{for } j = 1, \dots, m \{ \\ &\quad\quad R[i, j] = \min \begin{bmatrix} R[i - 1] + 1 & \text{(изтриване)} \\ R[i - 1, j - 1] & \text{(съответства)} \\ R[i - 1, j - 1] + 1 & \text{(замяна)} \\ R[i, j - 1] + 1 & \text{(вмъкване)} \end{bmatrix} \end{aligned}$$

$$B[i, j] = \begin{cases} 1, & \text{ако думата е изтрита} \\ 2, & \text{ако думата е вмъкната} \\ 3, & \text{ако думата съответства} \\ 4, & \text{ако думата е заменена} \end{cases}$$

Стъпка 3: Връщане назад и прекратяване

$$\text{Степен нагрешени думи} = 100\% * \frac{R(n, m)}{n}$$

Оптимален обратен път = $(S_1, S_2, \dots, 0)$ където $S_1 = B[n, m]$

$$S_t = \begin{cases} B[i-1, j], & \text{ако } S_{t-1} = 1 \\ B[i, j-1], & \text{ако } S_{t-1} = 2 \\ B[i-1, j-1], & \text{ако } S_{t-1} = 3, 4 \end{cases} \text{ за } t = 2, \dots, \text{докато } S_t = 0$$

Акустично моделиране. Извличане на характеристики от речевия сигнал и прилагане на адаптивни техники за намаляване на грешката

Преобразуване на свойствата на речевата вълна

За нормализиране на характеристиките на говорещия се използват:

- линеен дискриминантен анализ;
- картографиране, основано на невронни мрежи;
- изкривяване на честотния спектър при нормализиране на дължината на говорния тракт, чрез честотно мащабиране на Мел и билинейна трансформация.

За редуциране на външните различия между различни говорещи се използва коефициент на изкривяване, пресметнат за всеки от тях.

Избор на априорни единици при фонетичното моделиране

Фонетичните системи са свързани с конкретния език и неговите лингвистични правила. Важните елементи, които трябва да бъдат моделирани, са начина на представяне на акустичната тишина при липса на речеви сигнал и фонетичната информация за езика. Единиците, които се моделират трябва да бъдат:

- точни - да представят акустичните особености на фонемите в различните им употреби;
- обучени - да имат достатъчно данни за правилно оценяване на съответния елемент;
- породени - всяка нова дума да може да бъде получена като производна на предварително дефинирани елементи в системите за разпознаване на реч.

Сравнение на различните елементи

Независимо от това дали се използват контекстно зависими, или контекстно независими модели на думи, е необходимо дефинирането и автоматичното определяне на всички възможни фонологични вариации в конкретния език. Когато тези модели на думи са адекватно обучени, те предоставят най-добрите резултати в процеса на разпознаване. Следователно при разпознаване на малък речник, моделите от цели думи са много по-широко използвани, тъй като са точни и не е необходимо да бъдат обобщени и всеобхватни. Моделите на думи са точни при наличието на достатъчен набор от данни, като така биват обучени само върху ограничен набор от задачи, но така тяхното реално приложение е ограничено.

Компромисът между моделите, основани на думи и тези, основани на фонетика е, че при вторите се използва по-големи единици (пр. срички). Тези единици включват обособени групи, различаващи отделните фонетични единици (фонетични клъстери), които обхващат всички изменения на една фонетична единица, в зависимост от конкретиката на контекста. Центърът на клъстера е точка, независеща от контекста, докато началната и крайната точки са силно определени от контекста.

Контекстно зависими елементи

Използването на контекстно зависими елементи значително подобрява точността в процеса на разпознаване, при условие че има достатъчно данни за оценяване на тези контекстно-зависими параметри. Контекстно-зависимите фонемни са широко използвани за разпознаване на реч при голям речник от думи, тъй като значително подобряват тяхната точност и ефективност. При този метод от значение са съседните фонемни, намиращи се в ляво и в дясно.

Ако използваният фонетичен модел е контекстно зависим, качеството на разпознаване значително може да бъде подобро при условие, че има достатъчно данни за обучение и всички контекстно-зависими параметри са определени. Контекстно-зависимите фонемни са широко използвани в системите за разпознаване на реч при голям речник.

Съществува фонетичен модел, който взема под внимание както левите, така и десните съседни фонемни и това е трифонният модел. Ако две фонемни са едни и същи, но с различен ляв или десен контекст, те се разглеждат като различни трифони. Различните реализации на една фонема се наричат алофони. Трифоните са частен случай на алофоните. Моделиране на контекстно-зависими фонемни намиращи се в отделни думи е сложен процес.

Клъстеризиране на акустично-фонетични единици

Балансирането на способността за обучение и точността между фонетичния модел и модела, използващ думи, може да се обобщи чрез моделиране на подфонетични събития. В действителност резултатите, постигнати както с фонетичните, така и с подфонетичните данни, са сравнително еднакви. Това е ключовото предимство пред моделирането на цели думи. [49]

При подфонетичното моделиране се разглеждат състояния във фонетични скрити модели на Марков. След допълнително обобщаване на изходния клъстер, резултатите са разпределени според различните фонетични модели. По този начин всеки клъстер представлява набор от сходни Марковски състояния, наричен сенон. Така моделът от поддуми, който се състои от последователност от сенони, след тяхното клъстеризиране е завършен. Оптималният брой от сенони за една система се определя от наличната база за обучение и може да бъдат допълнен в последствие.

Адаптивни техники и намаляване на несъответствията

Разработени са редица техники за използването на адаптивни техники с цел намаляване на несъответствията. Може да се използва неинтрузивен обучителен процес, който през цялото време работи на фонов режим. При обучение без учител параметрите на модела могат непрекъснато да бъдат променяни, така че всички несъответствия се отстраняват. Системите, които транскрибират реч не в реално време могат да използват различни комбинации на техники за разпознаване.

Тъй като резултатите от разпознаването могат да бъдат непълни, съществува възможност за отклонения, ако нивото на грешката при разпознаване е високо.

Дори и нивото на грешка е ниско, адаптираните резултати не са толкова добри, колкото при обучение с учител. Чрез този вид обучение могат да се управляват широка гама от параметри:

- проверка на фоновия шум, когато дикторът не говори;
- регулиране на усилването на микрофона при нормален говорене на диктора;
- адаптиране на акустичните параметри чрез прочит на няколко изречения от диктора;
- промяна на декодиращи параметри, така че да се постигне най-добрата скорост без загуба на точност;
- динамично създаване на нови записвани изречения въз основа на модели за грешки, специфични за потребителя.

Максимална апостериорна оценка

Максималната апостериорна оценка (*Maximum a posteriori – MAP*) може ефективно да се справи с проблема, свързан с недостатъчни данни, тъй като може да се възползва от предварителната информация за съществуващите модели. Параметрите могат да бъдат коригирани при повторното обучение, така че ограничените нови обучаващи данни биха могли да променят параметрите на модела, ръководен от предварителните знания, и така да се компенсира резултатът от несъответствията. Подобряване на параметрите предотвратява големите отклонения, и освен това позволяват подобряване на крайните резултати в процеса на разпознаване.

Максимална вероятностна линейна регресия

Максималната вероятностна линейна регресия (*Maximum likelihood linear regression – MLLR*) отговаря на основните критерии за създаване на скрити модели на Марков, в случай, че броят на свободните параметри се запази. Тъй като параметрите на преобразуването могат да бъдат оценени при сравнително малко адаптирани данни, то максималната вероятностна линейна регресия е подходяща за ускоряване на адаптирането. Този метод е широко използван за адаптиране на модели, като при нови диктори, така при промяна на средата.

Сравнение на максимална вероятностна линейна регресия и максималната апостериорна оценка

Максималната вероятностна линейна регресия може да бъде комбинирана с максималната апостериорна оценка. Това гарантира, че с увеличаване на обема от данни за обучение ще има не само компактна максимална линейна регресионна преобразуваща функция за бързо адаптиране, но също така и директно преобразувани параметри на модела.

Клъстерни модели

Разпределението по отделни клъстери отчита разнообразни вариации спрямо различните фонетични класове. Освен това е възможно допълнително класифициране на подфонетично ниво.

Доверителни оценки

Един от повечето критични компоненти в практическите системи за разпознаване на реч е свързана с надеждността на доверителните оценки. В системите за разпознаване на реч, зависещи от диктора и тези, които не зависят от диктора, за определяне на доверителни оценки се използват записи на дикторите за премахване на думите, които могат да бъдат объркани. Това е критично в системите с обучение без учител на дикторите, позволявайки избирателно използване на резултатите от разпознаването, така че траскрипциите с по-малка сигурност и яснота могат да бъдат отхвърлени от процеса на адаптация.

Изкуствени невронни мрежи

Изкуствените невронни мрежи са модели на биологичните невронни мрежи.

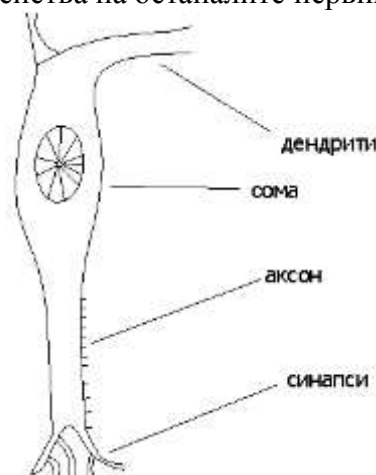
Невронната мрежа е система за паралелно обработване на информация, която има възможност да съхранява и използва експериментални данни и знания. Тя моделира дейността на своя биологичен еквивалент - мозъка в следните два аспекта:

- информацията се натрупва в мрежата чрез процес на обучение;
- силата на връзките между отделните възли (неврони) се моделира с тегла на съответните връзки, които се използват за съхранение на информацията.

Най-общо невронните мрежи се състоят от прости елементи за обработка на информация наречени неврони или възли. Невроните са свързани и теглата на връзките между тях определят силата на съответните връзки. Входната информация за всеки неврон е претеглената сума от сигналите от останалите неврони. Тази информация се акумулира в неврона, като изходния му сигнал се определя посредством активационна (предавателна) функция.

Биологичен неврон

Опростен модел на биологичен неврон е показан на фигура 4. Той се състои от синапси, дендрити, сома (тяло на клетката) и аксон. Дендритите и аксона са израстъци на клетката, чрез които тя е свързана с останалите клетки. Посредством дендритите, клетката получава сигнали от аксоните на другите клетки в мозъка. В синапсите се осъществява връзката и сигналите се преобразуват чрез сложни електрохимични процеси. Сигналите от синапсите се предават по дендритите към тялото на клетката, където се агрегират. Когато нивото на агрегирания сигнал надхвърли определена граница, се генерира сигнал по аксона на клетка и така тя въздейства на останалите нервни клетки.



Фиг. 4. Биологичен неврон

Абстрактен неврон

Изкуствените невронните мрежи съвсем не са точни и пълни модели на техните биологични аналози. Те са силно опростени и реализират само малък брой от техните добре изучени и изяснени структурни и функционални характеристики (фиг. 5).

В общия случай всеки неврон има много входове, които са модел на дендритите и един изход, който е модел на аксона. На входовете $x_1, x_2, \dots, x_m$ постъпват сигналите към неврона. Те могат да са външни сигнали или сигнали от изходите на други неврони. С всеки вход е свързано тегло $w_j, j = 1, 2, \dots, m$ (синаптично тегло), което моделира силата на връзката при предаване на сигнала посредством синапса, свързан със съответния дендрит. Агрегирането на входните сигнали в тялото на неврона се моделира със сума за пресмятане на изходния сигнал:

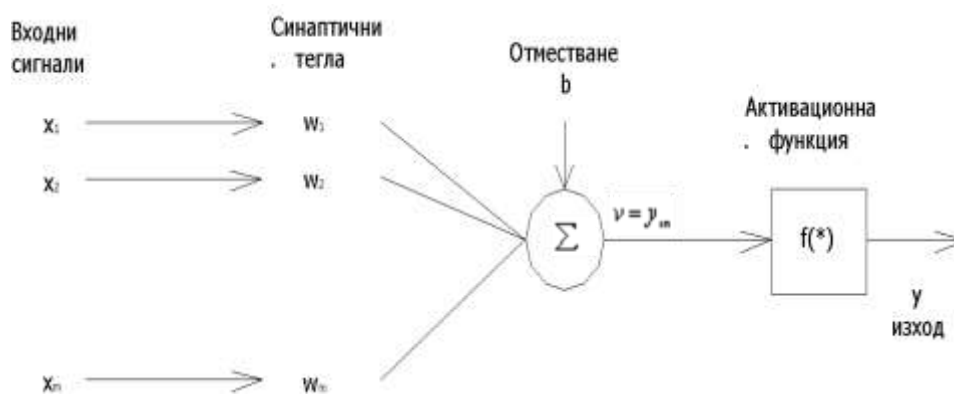
$$v = y_{in} = \sum_{j=1}^m w_j x_j + b,$$

където b е сигнал с постоянна стойност, наречен отместване.

Тялото на нервната клетка се моделира със суматора и блока на активационната функция. В общия случай активационната функция е нелинейна и изходният сигнал се изчислява по формулата:

$$y = f(y_{in}) = f\left(\sum_{j=1}^m w_j x_j + b\right).$$

В зависимост от начина на формулиране на активационната функция, изкуствените неврони могат да имат статично и динамично поведение. [50]

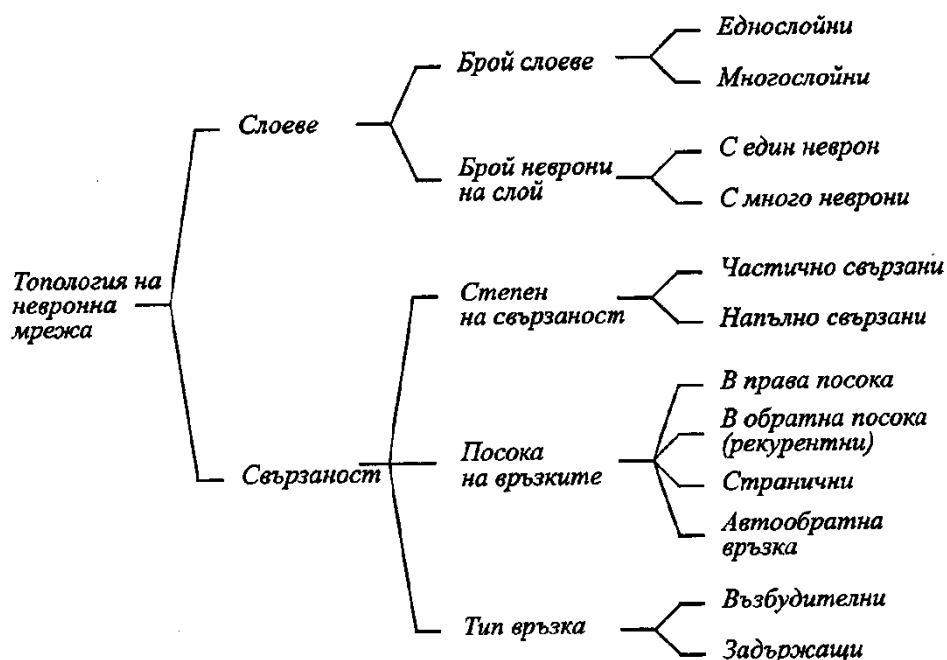


Фиг. 5. Модел на абстрактен неврон

Топология и Архитектури на изкуствените невронни мрежи

Топологията на невронните мрежи определя начина, по който множеството неврони в мрежата са подредени и свързани. Тя е най-важната характеристика на невронните мрежи, защото определя функционалните им особености и областите на приложението им. Признаците, които определят различните топологии са представени на фигура 6.

Броят слоеве и неврони в даден слой зависи от сложността на решаваната задача. Невроните могат да се свързват по различен начин и в зависимост от това да имат различно поведение спрямо информационния поток в мрежата. Когато даден неврон е свързан с малък брой други възли, мрежите се изграждат като частично свързани. Но в повечето приложения те се проектират като напълно свързани, а след трениране на мрежата несъществените връзки получават нулеви тегла, което е равносилно на отсъствие на връзка.



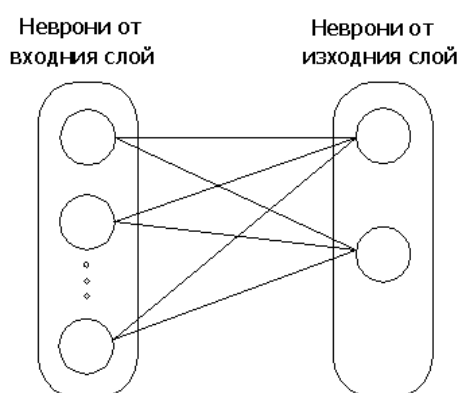
Фиг. 6. Топология на изкуствените невронни мрежи

Мрежите с *прави връзки* предават информация само в една посока – от входа към изхода. Ако има връзки в обратна посока, мрежите се наричат с *обратна връзка* или *рекурентни*. Възможно е да има и странични връзки между невроните в един слой. Тогава мрежите се наричат *странични*. Когато в мрежата се предвиди самите неврони да имат обратна връзка, те се наричат мрежи с *автообратна връзка*.

Връзките могат да бъдат *възбуждащи* или *поглъщащи*. Това се постига чрез задаване съответно на положителни или отрицателни стойности на теглата на съответната връзка. Например, при мрежите със състезателно обучение има както възбуждащи така и поглъщащи връзки. [51]

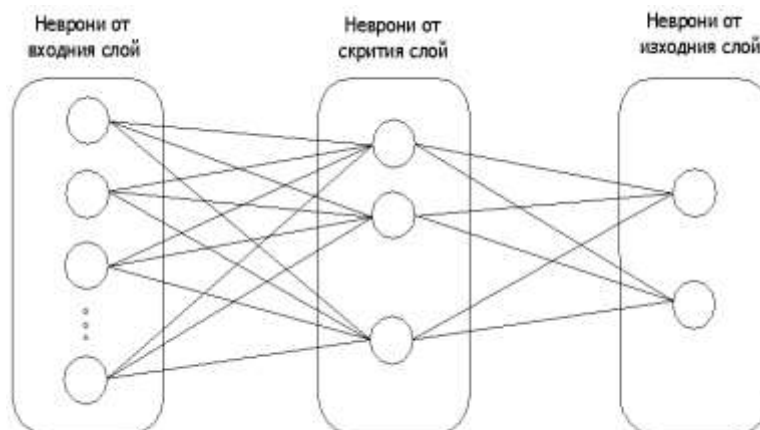
В приложенията най-често срещаните архитектури на невронните мрежи са:

- еднослойни мрежи с еднопосочно предаване на сигнала (фиг. 7.);



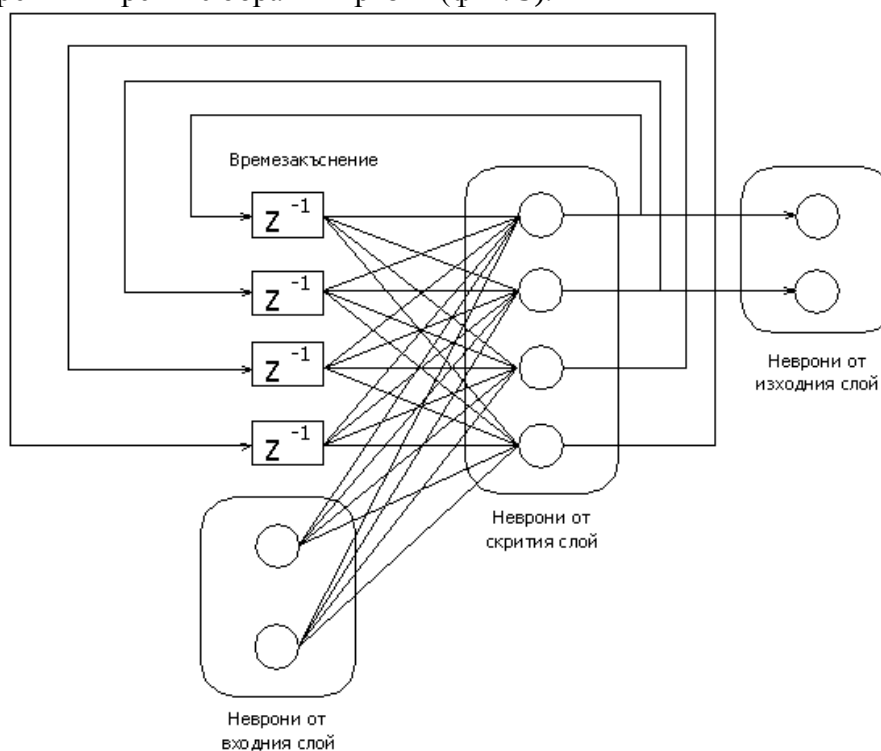
Фиг. 7. Еднослойна мрежа с еднопосочно предаване на сигнала

- многослойни мрежи с еднопосочно предаване на сигнала (фиг. 8.);



Фиг. 8. Многослойна мрежа с еднопосочно предаване на сигнала

- рекурентни мрежи с обратни връзки (фиг. 9).



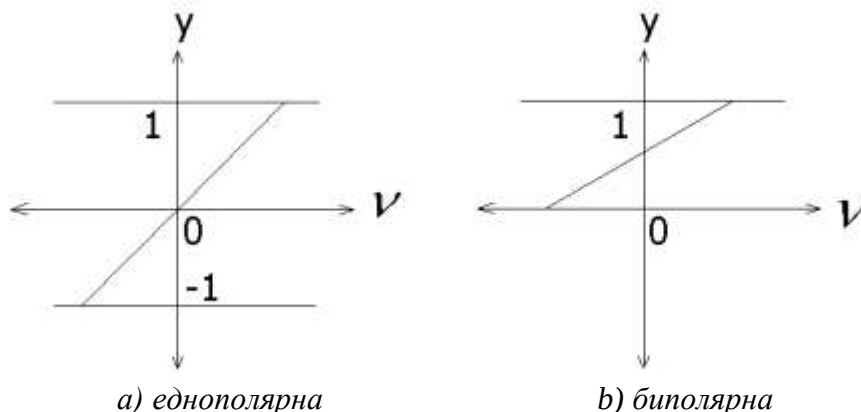
Фиг. 9. Рекурентна мрежа

Активационни функции

Направените изследвания са показали, че невронните мрежи работят устойчиво само при ограничени активационни функции, точно както и биологичните им еквиваленти формират изходен сигнал след като сумарното стимулиращо въздействие достигне определена граница. Ако изходният сигнал се изменя в интервала $[0; 1]$ те се наричат мрежи с еднополярни активационни функции, а изменението на стойностите е в интервала $[-1; 1]$ - мрежи с биполярни активационни функции). [50]

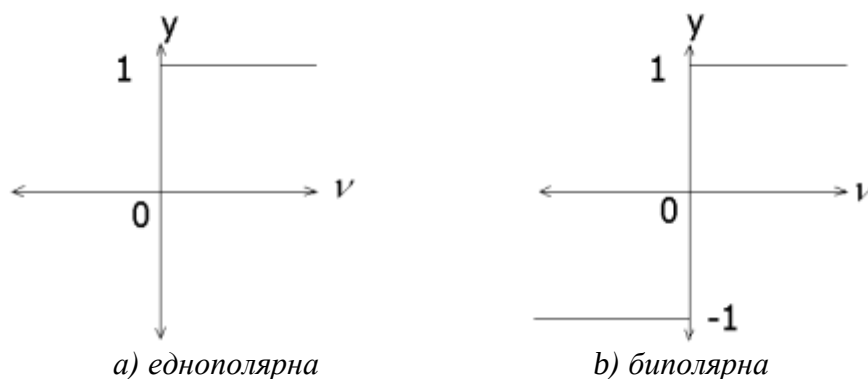
Две са основно използваните активационни функции :

- *линейна функция* – еднополярна (фиг. 10a) и биполярна (фиг. 10b);



Фиг. 10. Линейна активационна функция

- *прагова функция* - еднополярна (фиг. 11a) и биполярна (фиг. 11b). Праговата логическа функция (наричана още релейна, бинарна или сигнум функция) се използва за реализация на прости бинарни мрежови възли. Прилага се в еднополярен вариант (изходът има стойност 0 или 1) и двуполярен вариант (изходът приема стойност -1 или 1). Мрежите с бинарни неврони са най-лесни за хардуерна реализация, но често са много ограничени спрямо възможностите си за обучение [51];



Фиг. 11. Прагова функция

- *сигмоидална и радиално базисна функция*. На фигура 12a) е представена еднополярна сигмоидална функция, а на фигура 12b) – биполярна сигмоидална функция. Те се задават с уравнението

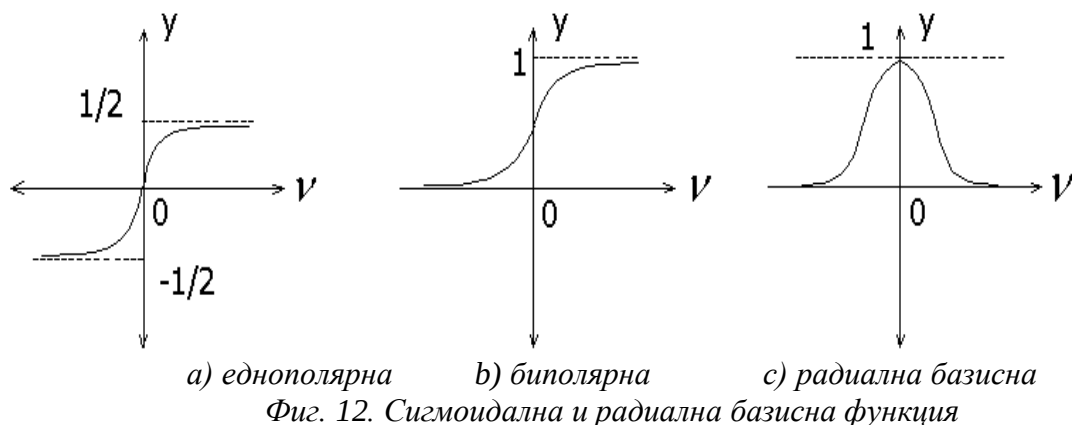
$$y = \frac{1}{1 + e^{-\alpha x}}$$

където α е константа, чрез която регулира ширината на прехода, като обикновено интервала на функционални стойности е $(0; 1)$, $(-1; 1)$ или $(-\frac{1}{2}; \frac{1}{2})$. Сигмоидални функции се използват при обучение с обратно разпространение на грешката.

Радиална базисна функция е представена на фигура 12c) и се задава с формулата:

$$y = e^{-\sum_{m=1}^n \frac{(x-c_m)^2}{2\sigma_m^2}},$$

където c_m задава центъра на активационната функция на m -тия неврон, σ_m - нейната изпъкналост. За разлика от останалите активационни функции, радиалната базисна функция придобива максималната си стойност единствено когато всички входни сигнали са със стойности близки до определената за тях централна стойност и е практически 0, когато стойностите се различават съществено от зададената централна. Невронни мрежи с радиално базисни функции се използват успешно в задачите за идентификация на обекти и управление. [51]



Обучение на невронни мрежи

Обучението на невронните мрежи е процес, при който чрез целенасочено изменение на тегловните коефициенти на връзките, мрежите придобиват желаните свойства или поведение. Подходите за обучение могат да се обединят в три групи. Всяка от тях наподобява определен аспект от поведенческото обучение в биологичните системи.

1. Обучение с учител

При *обучение с учител (supervised learning)* началните стойности на теглата са зададени предварително и така се подава вектор с входни данни на входа на мрежата. Получената реакция на изхода се сравнява с предварително зададена и се пресмята грешката. Нейната стойност е критерий за промяна на тегловните коефициенти. Този процес продължава до минимизиране на грешката, затова обучението с учител е известно и под наименованието *обучение с минимизиране на грешката*. Прилагането на метода изисква задаване на серии от входни данни - образци и желани изходни сигнали, които се наричат тренировъчни комплекти. Чрез тях системата се обучава да реагира по желания начин. При мрежите, използващи обучение с учител се разграничават два етапа – етап на трениране, при който теглата се адаптират за да бъде научено желано поведение и на опериране, при който намерените на първия етап тегловни коефициенти са фиксирани и мрежата работи с тях при различни входни данни. Този подход за обучение широко се използва при невронни мрежи с прави връзки и някои рекурентни мрежи.

2. Обучение без учител

При *обучение без учител (unsupervised learning)* теглото на дадена връзка се определя в зависимост от корелацията между активациите на свързаните чрез нея два неврона, откъдето произлиза и другото название - *корелационно обучение*. Първият вариант на такова обучение е предложен от Хеб и извършва актуализация на стойността на всеки тегловен коефициент в мрежата чрез добавяне към текущата му стойност на нов член, пропорционален на произведението от стойностите на активация на двата свързани неврона. Характерно за това обучение е, че то не изисква тренировъчни комплекти и не съществува разграничение между етапите на обучение и опериране на мрежата.

Обучението без учител намира широко приложение при задачи, които изискват асоциативна памет.

3. Обучение с поощрение

При *обучение с поощрение (reinforcement learning)*, подобно на обучението с учител, трябва да има примери с образци от входни данни, но не е необходимо да се предоставят образци с желаните изходни реакции. Въвежда се обща мярка за адекватност на получените резултати, която води мрежата към желаното поведение. Тази мярка се нарича усилващ или поощряващ сигнал, който се подава като обратна връзка в мрежата, за да се поощрят коректните и съответно накажат некоректните поведения. Това става чрез увеличаване на онези тегла, които са допринесли за доброто поведение и намаляване предизвикателите лош резултат. [51]

Приложение на невронните мрежи

При алгоритмичния подход за решаване на различни задачи в компютърните науки се прави абстрактен модел на разглежданата задача, след като се разработва алгоритъм за решаването и, и накрая алгоритъмът се реализира с компютърна програма. Алгоритъмът е ефективен метод, който при даден списък от коректно дефинирани команди за изпълнение на задача и дадено начално състояние, преминава през коректно дефинирана поредица от последователни състояния и завършва в едно крайно състояние. Преходът между състоянията не е задължително да е детерминиран, тъй като някои алгоритми, известни като вероятностни алгоритми, съдържат елемент на случайност. В компютърните науки алгоритмите се използват за пресмятане, обработка на данни и други. [52]

За разлика от алгоритмичния подход, използването на невронни мрежи се базира на обучение чрез примери, с което се прави опит за моделиране на биологичния им еквивалент - мозъка. По принцип с невронни мрежи може да бъде изчислена всяка изчислима функция, т.е. с невронни мрежи може да се прави всичко, което се прави със стандартните цифрови компютри. За разлика от алгоритмичния подход обаче, невронните мрежи използват принципа на паралелна обработка на информацията и поради това те могат да бъдат използвани успешно при задачи за обработване на данни (сигнали, изображения и др.) в реално време, когато се цели постигане на по-голямо бързодействие.

Невронните мрежи работят на принципа на черната кутия, като се използва информация от разглеждания процес или явление, с цел обучаване при настройване на теглатата на мрежата. По тази причина те са удобно средство за моделиране на процеси и явления, за които може да бъде събрана много входно-изходна информация (от измерване и др.), но трудно могат да бъдат описани аналитично.

Невронните мрежи са ефективни при решаване задачи за класификация и апроксимация. При наличие на достатъчно данни и компютърни ресурси при обучението, почти всяко изображение между две векторни пространства може да бъде апроксимирано с желана точност посредством многослойна невронна мрежа с едноръчно предаване на сигнала. Невронните мрежи намират широко приложение в почти всички области на науката и живота. Те се използват в:

- инженерната дейност за обработка на сигнали и изображения, за филтрация, при разпознаване на образи, при моделиране, идентификация и управление на сложни многосвързани обекти и др.;
- приложната математика за решаване на различни видове задачи като апроксимация, оптимизация с или без ограничения, решаване на различни видове уравнения и системи уравнения и др.;

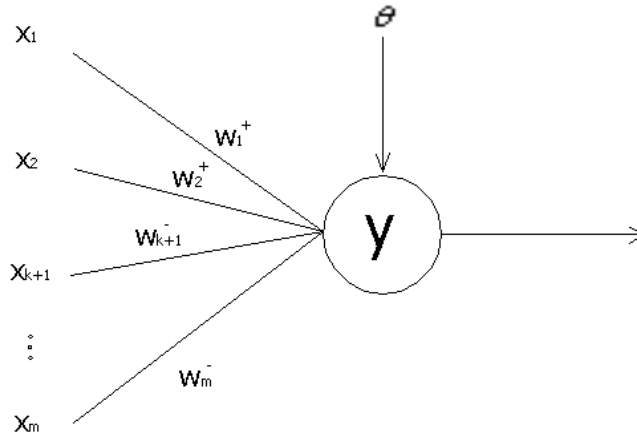
- статистиката за нелинейна регресия и класификация, прогнозиране на времеви редове и др.;
- когнитивните науки за описание и моделиране на мисленето и съзнанието;
- икономиката и финансите за изграждане на сложни икономически модели, за моделиране и прогнозиране на финансови пазари и много други. [50]

Видове невронни мрежи

Еднослойни невронни мрежи с прави връзки

1. Неврон на Маккалък-Пит

Невронът на Маккалък-Пит, предложен през 1943 г., е първото използване на невронна мрежа за изчисления (фиг. 13). Той се характеризира с изчислителна универсалност, тъй като всяка логическа функция може да бъде изчислена с мрежа от неврони на Маккалък-Пит. Същевременно всяка крайна последователност от дискретни действия може да бъде стимулирана с рекурентни невронни мрежи от такива неврони.



Фиг. 13. Неврон на Маккалък-Пит

При този вид невронна мрежа теглата са фиксирани, т.е. не се настройват с обучение. Входните сигнали $x_1, x_2, \dots, x_m$ и изходните сигнали y на невроните са бинарни - $x_i \in \{0, 1\}$, $i = 1, \dots, m$ и $y \in \{0, 1\}$. Входните сигнали биват два вида – възбудителни и забранителни. Теглата, свързани със възбудителните входни сигнали $x_1, x_2, \dots, x_k$ са положителни, а всеки ненулев входен сигнал на някой от забранителните входове $x_{k+1}, x_{k+2}, \dots, x_m$ генерира нулев изходен сигнал. Изходен сигнал 1 се получава, когато сумата от теглата на възбудителните входни сигнали е по-голяма от предварително зададен праг θ и няма сигнал на никой от забранителните входове [50]:

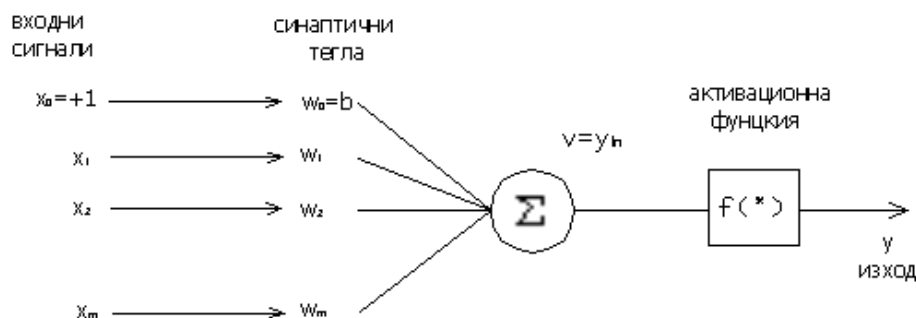
$$y = \begin{cases} 1, & \text{когато } \sum_{j=1}^k w_j^+ x_j \geq \theta \text{ или } x_i = 0, \quad i = k+1, \dots, m \\ 0, & \text{когато } \sum_{j=1}^k w_j^+ x_j < \theta \text{ или } x_i = 1, \quad i = k+1, \dots, m \end{cases}$$

За правилната работа на неврона трябва да е изпълнено условието:

$$\sum_{j=1}^k w_j^+ - w_j^- < \theta, \forall i = k+1, \dots, m$$

2. Адаптивен линеен елемент

Адаптивният линеен елемент (*ADaptive LINEar Element=ADALINE*) е предложен през 1962 г. от Бернард Уидроу. Той представлява единичен неврон с линейна активационна функция и настройващи се тегла (фиг.14). Алгоритъмът на Уидроу-Хоф за настройка на теглата използва метода на най-малките квадрати и е пример за използване на обучение при настройката. Той е в основа за различни видове адаптивна филтрация, използвана при обработка на сигнали.



Фиг. 14 Линейна активационна функция

За обучението на мрежите от тип *ADALINE* е необходимо да се разполага с обучаваща извадка с обем p , представляваща набори от входни сигнали и желан изход на мрежата за всеки от тях. При подаване на всеки k -ти комплект входни сигнали се определя изходът от невронната мрежа и грешката между него и еталонния изход при зададената комбинация на входа. Целта на обучението е теглата $w = [w_1 \ w_2 \ \dots \ w_m]^T$ да бъдат настроени така, че да бъде минимизирана грешката $\sum_{i=1}^p e_i$ за цялата обучаваща съвкупност. В различни варианти на обучаващите алгоритми промяната на теглата се прави след всеки компонент от обучаващата серия или след всяка епоха. Във всички случаи промяната на теглата е пропорционална на съответния входен сигнал, грешката и параметър η , наречен скорост на обучение.

Мрежите *ADALINE* принадлежат към общия клас алгоритми, наречени адаптивни линейни филтри. Те намират приложение в много области - за проектиране на изравняващи филтри при високоскоростни модеми, за адаптивни ехокомпенсатори при телефонни разговори на големи разстояния и сателитни комуникации, както и за предсказване на сигнали. *ADALINE* се прилагат и в медицината - за подтискане на шума от биене на сърцето на майката при ембрионална електрокардиография.

3. Перцептрон

Перцептронът е предложен през 1957 г. Розенблат. За разлика от адаптивния линеен елемент перцептронът е нелинеен неврон. Той използва релейна активационна функция (фиг. 15).

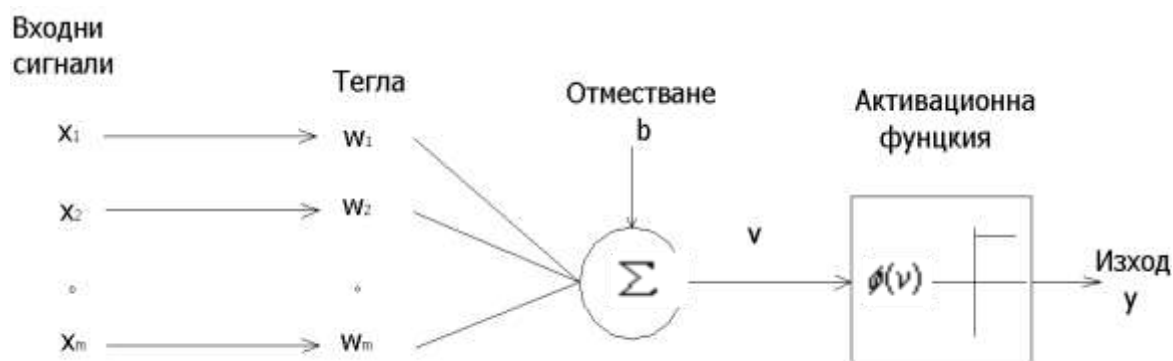
Изходът на суматора се изчислява като:

$$v = \sum_{j=1}^m w_j x_j + b = \sum_{j=0}^m w_j x_j$$

където $w_0 = b$, а x_0 е сигнал с постоянна стойност 1. Изходният сигнал на перцептрона се определя с формулата:

$$y = \phi(v) = \begin{cases} 1, & \text{ако } v \geq \theta \\ 0, & \text{ако } v < \theta \end{cases}$$

където θ е праг на активация на перцептрона.



Фиг. 15 Релейна активационна функция

При перцептрона се използва обучение с учител. При този тип обучение критерият за обучението се задава с комплект примерни входно-изходни последователности за желаната работа на мрежата от типа

$$\{p_1, d_1\}, \{p_2, d_2\}, \dots, \{p_Q, d_Q\}$$

където p_k се подават на входа на мрежата, а d_k е съответната правилна стойност на изхода на мрежата. След сравняване на d_k с получения резултат се стартира алгоритъм за обучение, посредством който се настройват теглата и отместванията на мрежата така, че получаваната стойност на изхода на мрежата да се доближава до желаната.

Обучението се извършва по следното правило:

- ако $d = 0$ и $\phi(w^T x) = 1$, то $w_{new} = w_{old} - x$,
- ако $d = 1$ и $\phi(w^T x) = 0$, то $w_{new} = w_{old} + x$,
- ако $d = \phi(w^T x) = 1$, то $w_{new} = w_{old}$,

където с w_{new} се означава новата, а с w_{old} - старата стойност на тегловия коефициент от преходната интерация; с x се означава съответната стойност на p_k , подадена на входа и с d съответното d_k . Функцията $\phi(w^T x)$ приема в случая стойност 0 или 1, т.е.

$$y = \phi(v) = \begin{cases} 1, & v \geq \theta \\ 0, & v < \theta \end{cases}$$

Ако бъде въведена грешка $e = d - \phi(w^T x)$, правилото за обучение може да бъде представено в следния компактен вид: $w_{new} = w_{old} + e \cdot x$.

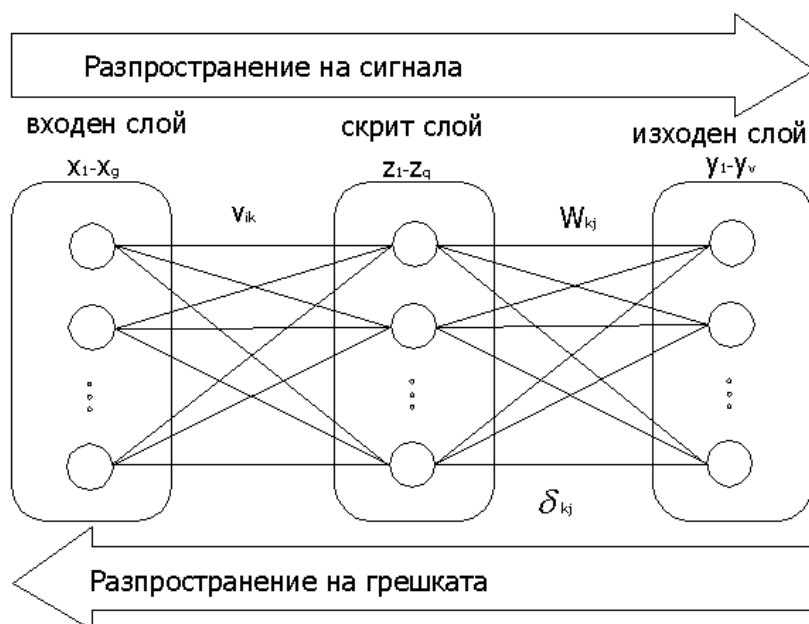
Правилото може да се обобщи за един слой от K на брой перцептрони, използващи едни и същи входни сигнали $x = [x_1 \ x_2 \ \dots \ x_m]^T$. В този случай формулите за настройка на вектора с теглата w_s и отместване b_s , свързани с s -тия перцептрон са равни на: $w_{s_new} = w_{s_old} + e_s \cdot x$, $b_{s_new} = b_{s_old} + e_s$, където $e_s = d_s - \phi(w_s^T x + b_s)$.

Многослойни невронни мрежи с прави връзки

1. Многослойни невронни мрежи с обратно разпространение на грешката

На практика най-често използваните многослойни мрежи са двуслойни, защото е доказано, че при достатъчно неврони в междинния слой, всяка двуслойна мрежа може да реализира функциите на мрежа с повече слоеве. Архитектурата на двуслойна невронна мрежа с еднопосочно предаване на сигнала, използваща метода на обратно разпространение на грешката е показана на фигура 16.

Всички алгоритми за обучение на многослойните невронни мрежи се основават на алгоритъма с обратно разпространение на грешката, който е известен още като обобщено делта правило. В основата си той представлява градиентен метод, целящ минимизиране на грешката на изхода на мрежата.



Фиг. 16. Невронна мрежа с еднопосочно предаване на сигнала, използваща метода на обратно разпространение на грешката

Обучението на многослойните невронните мрежи преминава през три етапа.

По време на първия, наречен *право разпространение*, всеки входен неврон получава входен сигнал и го изпраща на всеки от невроните от скрития слой. Невроните в скрития слой формират своя изходен сигнал и го изпращат на невроните от изходния слой. В края на процеса изходните неврони изчисляват своите изходни сигнали и по този начин формират реакцията на мрежата на подаденото входно въздействие.

На следващия етап от обучаващия процес, реакцията на всеки изходен неврон се сравнява с желаната реакция с цел определяне на грешката. На нейна основа се определя коефициентът δ_k за всеки неврон в изходния слой. Коефициентът δ_k се използва, за да се разпредели грешката от изходния слой към предходния. По-късно тя се използва за преизчисляване на теглата между скрития и изходния слой. По подобен начин се изчислява и δ_j за скрития слой. Не е необходимо разпространението на грешката назад към входните неврони, но тя се използва за обновяване на стойността на тегловните коефициенти между входа и скрития слой.

В следващия трети етап, след като всички коефициенти δ бъдат определени, теглата на всички слоеве на мрежата се обновяват едновременно.

Активационната функция на мрежата с обратно разпространение на грешката трябва да има няколко важни свойства – да бъде непрекъсната, диференцуема и монотонно растваща. Друго условие е функцията да има крайни максимум и минимум, които да се достигат асимптотично. Най-често използваната функция, отговаряща на горните условия, е сигмоидалната.

Многослойните мрежи, използващи обучение с обратно разпространение на грешката имат няколко съществени предимства:

- могат да моделират всяка непрекъсната функция;
- нуждаят се само от подходяща обучаваща извадка;
- не изискват предварително познаване на проблемите и затова са приложими за решаване на лошо структурирани задачи;
- устойчиви са на шум и липса на данни от обучаващата извадка.

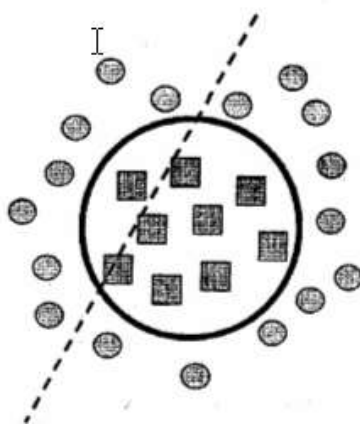
Наред с това този вид невронни мрежи притежават и някои недостатъци:

- обучението може да изисква голям обем ресурси;
- при обучението на мрежата е възможно достигане на локален минимум, който да е разичен от глобалния минимум.

За преодоляване на последния проблем са предложени редица модификации на базовия обучаващ алгоритъм. Един ефективен подход е ново стартиране на обучението с променени начални тегла и/или разместване на обучаващите комплекти в извадката.

Многослойни невронни мрежи с радиални базисни функции

Идеята при въвеждане на тези мрежи е свързана с използването на линейни разделящи правила при класификация на образи. За двумерния пример от фигура 17, двата вида образи са разделими с нелинейна крива (показаната окръжност), но са линейно неразделими, т.е. не съществува права която да ги разделя.



Фиг. 17. Двумерен пример

Невронните мрежи, способни да се справят с този проблем, са мрежи с радиална базисна функция (фиг. 18). Те са двуслойни мрежи, чиито неврони в скрития слой имат нелинейна активационна функция, а невроните в изходния слой – линейна активационна функция.

Трите основни параметъра, които влияят на качеството на мрежата с радиални базисни функции, са:

- центрове на радиално-базисните функции.
- радиусите на радиално-базисните функции.
- теглата на връзките в изходния слой.

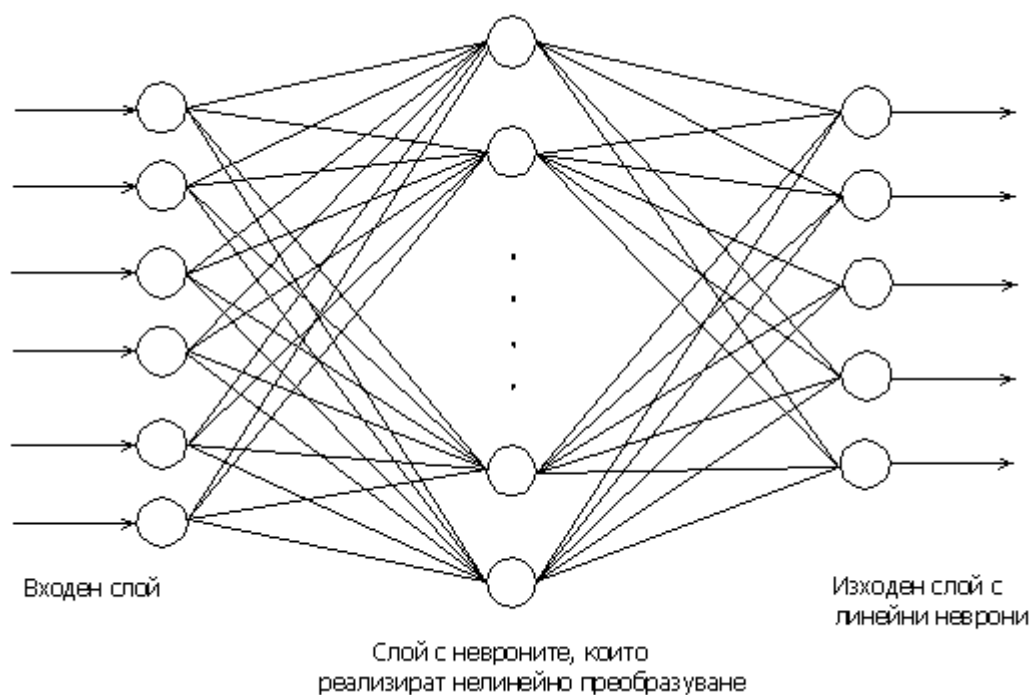
При обучение на тези мрежи обикновено центровете и радиусите на радиално-базисните функции се определят без учител, а за теглата се използва обучение с учител.

Радиусите на функциите се изчисляват по формулата:

$$\sigma = \frac{d_{max}}{\sqrt{2m}},$$

където d_{max} е максималното разстояние между два неврона, а m е броят на невроните в скрития слой.

Предимствата на мрежите с радиално базисни функции пред стандартните мрежи, използващи алгоритъм с обратно разпространение на грешката, са в това, че те имат по-бързо обучение и по-сигурно достигане на глобален минимум на грешката. Недостатък на тези мрежи е, че много често изискват по-голям брой неврони в скрития слой. [50]



Фиг. 18. Мрежа с радиални базисни функции

Рекурентни невронни мрежи

Невронните мрежи, в които информацията се пренася само в една посока – последователно от входния, през скритите слоеве (ако има такива) до изходния слой се наричат невронни мрежи с прави връзки. Съществуват мрежи, при които информация се движи и в обратна посока – към текущия или към предходни слоеве на мрежата. Тези мрежи се наричат *рекурентни мрежи*. Когато на една рекурентна мрежа се подаде входен вектор, не се получава директно изходен вектор след краен брой стъпки, а се предизвиква кръгов процес или резонанс на нервната активност в мрежата. В резултат на това едни и същи слоеве се активират многократно. Ако мрежата резонира известно време и след това се установи в устойчиво състояние, тя се нарича собственоустойчива. Целта на тренирането на рекурентните мрежи е да се намерят такива тегла на връзките, които да гарантират собствена устойчивост на мрежата при стабилизирани стойности на изхода, съответстващи с еталонни, зададени за обучаващите входни серии. [49]

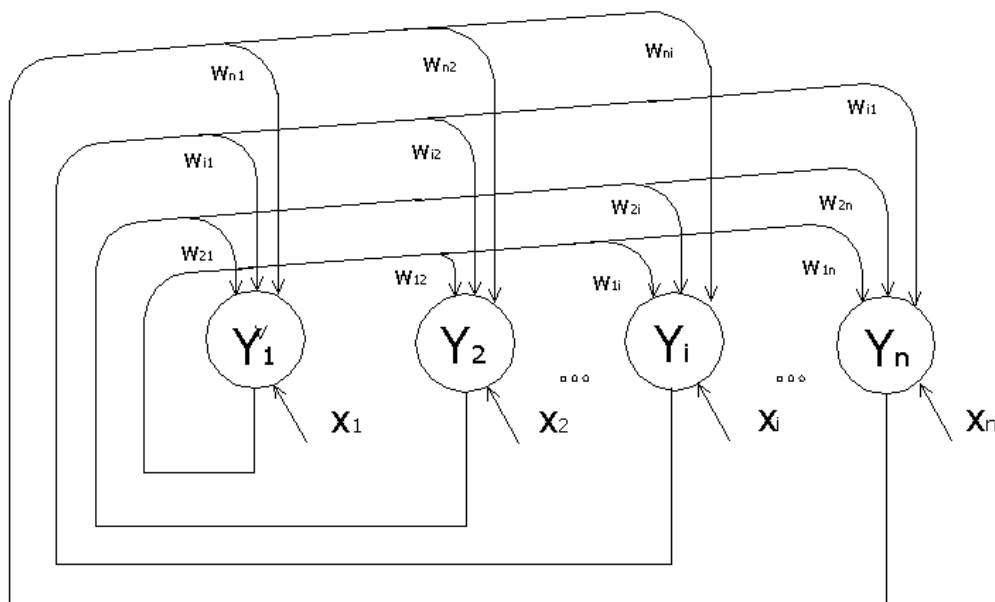
Съществуват различни архитектури на рекурентни невронни мрежи. Едни от най-разпространените са предложените от Хопфийлд, Елман и Джордан.

1. Невронни мрежи на Хопфийлд

Невронните мрежи на Хопфийлд са еднослойни мрежи, за които всеки неврон е свързан с всички останали. Те биват два вида: дискретни и непрекъснати и могат да бъдат използвани не само като асоциативни памети, а също така за решаване на сложни комбинаторни задачи.

Архитектурата на дискретна мрежа на Хопфийлд е показана на фигура 19. Теглата в мрежата са съответно $w_{ij} = w_{ji}$ и $w_{ii} = 0$, т.е. матрицата с теглата на мрежата е симетрична с нулеви елементи по главния диагонал. Входните сигнали $x_1 x_2 \dots x_n$ не са задължителни. В някои случаи входни сигнали може да има, а в други случаи може да няма, като входове за невронната мрежа са началните състояния на изходните сигнали на невроните, а изходи са изходните сигнали на невроните след съответния брой

интерации. В тези случаи изходните сигнали на невроните, които определят състоянието на мрежата на всяка интеракция, служат съответно за вход и изход на мрежата. Активационните функции на невроните са релейни, като в зависимост от данните (бинарни или биполярни) стойностите им са в $[0 ; 1]$ или $[-1 ; 1]$.



Фиг. 19. Невронни мрежи на Hopfield

Матрицата с теглата на мрежата се изчислява като се използва правилото на Хеб, като диагоналните ѝ елементи се нулират.

При биполярни данни:

$$w_{ij} = \sum_p S_i(p)S_j(p), \quad i \neq j$$

$$w_{ii} = 0$$

а при бинарни данни:

$$w_{ij} = \sum_p (2S_i(p) - 1)(S_j(p) - 1), \quad i \neq j$$

$$w_{ii} = 0$$

като в случая при конструиране на матрицата с теглата всеки бинарен вектор се преобразува в биполярен.

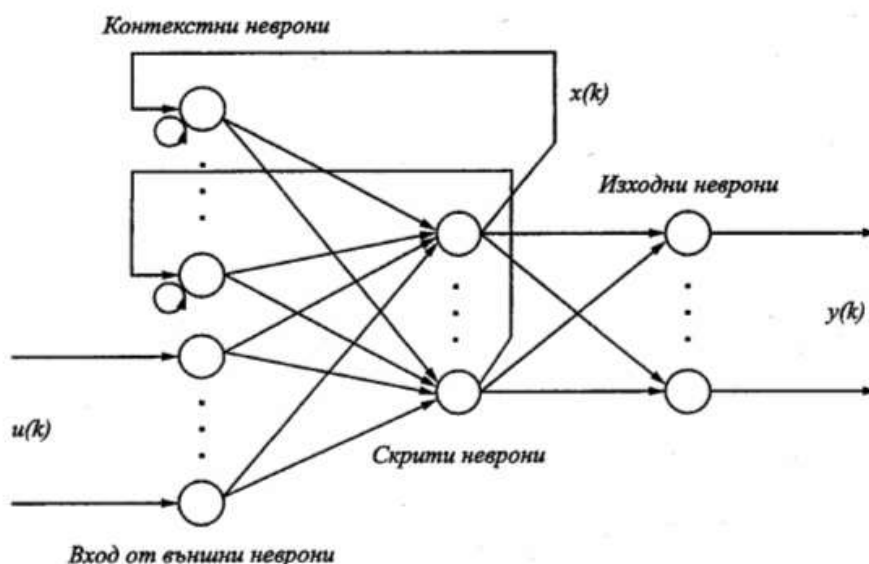
Най-важният момент при реализацията на интерациите е асинхронната промяна на изходните сигнали на невроните. Това означава, че на всяка стъпка се променя изходният сигнал само за един неврон, като този неврон се избира случайно. Вероятността за това произволен неврон да бъде избран на дадена стъпка е равна. Получената нова стойност на изхода на неврона се използва при промяната на изходния сигнал за следващите неврони. Този подход гарантира сходимостта на промяната на изходните сигнали на невроните на мрежата. [50]

2. Невронни мрежи на Елман

На фигура 20 е показана архитектурата на мрежата на Елман. Тя е трислойна. Първият слой се състои от две различни групи неврони – външни входни неврони и вътрешни входни неврони, наричани още контекстни възли. На входовете на контекстните възли постъпват изходите от неврони, разположени в скрития слой. Вътрешните входни

неврони се наричат и неврони на паметта, тъй като съхраняват стойността на изхода на неврон от скрития слой в предходен времеви такт. Вторият слой е скрит слой. На входовете на неговите неврони постъпват изходите на външните и вътрешните входни неврони.

Теглата на рекурентните връзки не се променят, като в общия случай се приемат за равни на 1. На настройване подлежат само теглата на правите връзки. При начална инициализация на мрежата на тях се присвояват произволни малки начални стойности, а на изходите от контекстните неврони – стойност 0,5 при използване на сигмоидна функция на активация.

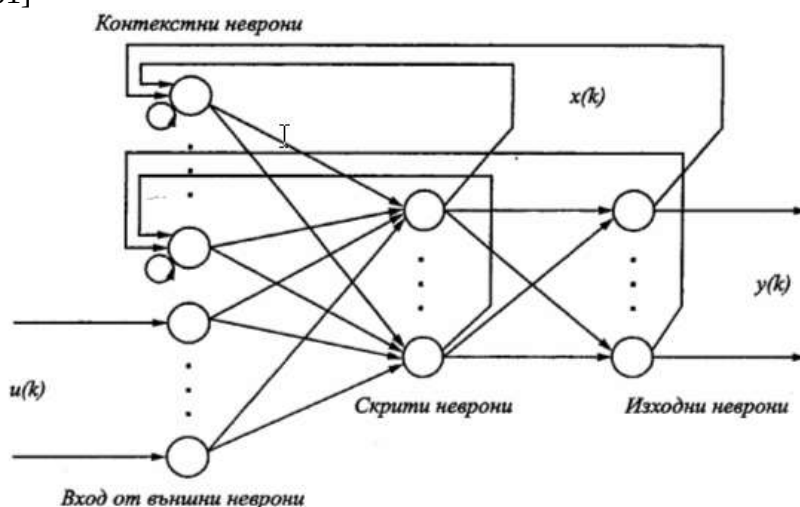


Фиг. 20 Невронни мрежи на Елман

3. Невронни мрежи на Джордан

Мрежата на Джордан (фиг. 21.) също е трислойна и с автообратни връзки при контекстните неврони, но обратните връзки са от изходния към контекстния слой. Наличието на обратни връзки и от скрития слой, подобно на мрежата на Елман позволяват чрез нея да се си моделират произволни динамични системи.

Мрежите на Джордан могат да бъдат тренирани чрез стандартния алгоритъм с обратно разпространение на грешката, като стойностите на теглата на обратните връзки са фиксирани. [51]



Фиг. 21. Невронни мрежи на Джордан

V. ЗАКЛЮЧЕНИЕ

Развитието на технологиите за гласово разпознаване започва през 60-те години на XX век с малък набор единични думи (10-100), които биват разпознавани на базата на акустично-фонетичните свойства на речевите звуци. Ключовите технологии, развили се през този ранен период са филтър анализи, методи за нормализация по време и е сожено началото на сложните методологии за динамично програмиране. През 70-те години се осъществява разпознаването на средно голяма лексика (100-1000 думи), като се използват шаблонно базирани методи за разпознаване на мостри. Основните технологии през този период са модели за разпознаване на шаблони, въвеждането на LPC методите за спектрално представяне, методи за създаване на клъстер от шаблони за независими от говорещия разпознаватели, и въвеждането на методите за динамично програмиране за решаване на свързани задачи за разпознаване на думи. През 80-те години се работи със системи за гласово разпознаване, използващи голяма лексика (>1000 думи), основавайки се на статистически методи, с голям набор от мрежи за справяне с езиковите структури. Ключовите технологии през този период са скритите модели на Марков и стохастичния езиков модел, които заедно са в основата на мощни нови методи за ефективно и високопроизводително решение на всяка задача, свързана с непрекъснато речево разпознаване. През 90-те години се въвеждат системи за лексика с неограничени езикови модели и ограничените по задачи синтактични модели за непрекъснато речево разпознаване и разбиране. Ключовите технологии по това време са методите за стохастично разбиране на езика, статистическо обучение на акустични и езикови модели, и въвеждането в трансформаторните системи от крайни състояния, като и методите за тяхното детерминиране и минимизиране за ефикасно внедряване в големите системи за разпознаване на говор.

В следващите десетилетия се въвеждат огромни от лексикална гледна точка системи с пълни семантични модели, интегрирани с речево-говорни синтезирани системи и много-модални входи (клавиатури, мишки и т.н.). Тези системи осъществяват речеви диалог с набор от входни и изходни модалности за лесна употреба и гъвкавост при работа с разнообразни среди, където речта може да не е най-подходящото средство тъй както други входно-изходни модалности. По време на този период се зараждат високоефективни естествени конкативни системи за синтезиране на речта, употребата на машинното обучение за подобряването на разбирането на реч и диалога, и се въвеждат смесено-инициативните диалогови системи, които предоставят потребителски контрол при необходимост.

Основно предизвикателството все още е задачата за проектиране на машина, която реално да функционира като интелигентно човешко същество. Постиженията до днешна дата са само началото и ще са нужни още много изследвания, докато при машините бъде постигната производителност в реално време, която съперничи на тази на човека.

References

- [1] Davis K., R. Biddulph, S. Balashek, "Automatic Recognition of Spoken Digits," *The Journal of the Acoustical Society of America*, vol. 24, no. 6, pp. 627-642, 1952.
- [2] Olson H., H. Belar, "Phonetic Typewriter," *The Journal of the Acoustical Society of America*, vol. 28, no. 6, pp. 1072-1081, 1956.
- [3] Forgie J., C. Forgie, "Results Obtained from a Vowel Recognition Computer," *The Journal of the Acoustical Society of America*, vol. 31, no. 11, pp. 1480-1489, 1959.

- [4] Suzuki J., K. Nakata, “Recognition of Japanese Vowels—Preliminary to the Recognition of Speech,” *J. Radio Res. Lab*, vol. 37, no. 8, pp. 193-212, 1961.
- [5] Sakai T., S. Doshita, “The Phonetic Typewriter,” *The Journal of the Acoustical Society of America*, vol. 33, no. 11, 1961.
- [6] Nagata K., Y. Kato, S. Chiba, „Spoken Digit Recognizer for Japanese Language,” *NEC Res. Develop*, № 6, 1963.
- [7] Denes P., “The Design and Operation of the Mechanical Speech Recognizer at University College London,” *British Institution of Radio Engineers*, vol. 19, no. 4, pp. 211-229, 1959.
- [8] Martin T., A. Nelson, X. Zadell, “Speech Recognition by Feature Abstraction,” Tech. Report AL-TDR-64-176, Air Force Avionics Lab, 1964.
- [9] Vintsyuk T., “Speech Discrimination by Dynamic Programming,” *Kibernetika*, vol. 4, no. 2, pp. 81-88, 1968.
- [10] Sakoe H., S. Chiba, “Dynamic Programming Algorithm Quantization for Spoken Word,” *Speech and Signal Proc.*, vol. 26, no. 1, pp. 43-49, 1978.
- [11] Viterbi A., “Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm,” *IEEE Trans. Informaiton Theory*, vol. 13, pp. 260-269, 1967.
- [12] Atal B., S. Hanauer, “Speech Analysis and Synthesis by Linear Prediction of the Speech Wave,” *J. Acoust. Soc. Am.*, vol. 50, no. 2, pp. 637-655, 1971.
- [13] Itakura F., S. Saito, “A Statistical Method for Estimation of Speech Spectral Density and Formant Frequencies,” *Electronics and Communications in Japan*, vol. 53, pp. 36-43, 1970.
- [14] Itajura F., “Minimum Prediction Residual Principle Applied to Speech Recognition,” *IEEE Trans. Acoustics, Speech and Signal Proc*, vol. 23, pp. 57-72, 1975.
- [15] Rabiner L., S. Levinson, A. Rosenberg, J. Wilpon, “Speaker Independent Recognition of Isolated Words Using Clustering Techniques,” *IEEE Trans. Acoustics, Speech and Signal Proc.*, vol. 27, pp. 336-349, 1979.
- [16] Lowerre B., “The HARPY Speech Understanding System,” *Trends in Speech Recognition, Speech Science Publications, 1986, reprinted in Readings in Speech Recognition*, pp. 576-586, 1990.
- [17] Mohri M., “Finite-State Transducers in Language and Speech Processing,” *Computational Linguistics*, vol. 23, no. 2, pp. 269-312, 1997.
- [18] Klatt D., “Review of the DARPA Speech Understanding Project (1),” *J. Acoust. Soc. Am.*, vol. 62, pp. 1345-1366, 1977.
- [19] Jelinek F., R. Bahl, R. Mercer, “Design of a Linguistic Statistical Decoder for the Recognition of Continuous Speech,” *IEEE Trans. On Information Theory*, vol. 21, pp. 250-256, 1975.
- [20] Shannon C., “A Mathematical Theory of Communication,” *Bell System Technical Journal*, vol. 27, pp. 379-423, 623-656, 1948.
- [21] Juang B., S. Levinson, M. Sondhi, “Maximum Likelihood Estimation for Multivariate Mixture Observations of Markov Chains,” *IEEE Trans. Information Theory*, vol. 32, no. 2, pp. 307-309, 1986.
- [22] Juang B., “Maximum Likelihood Estimation for Mixture Multivariate Stochastic Observations of Markov Chains,” *AT&T Tech. J*, vol. 64, no. 6, pp. 1235-1249, 1985.
- [23] Lee C., L. Rabiner, R. Pieraccini, J. Wilpon, “Acoustic modeling for large vocabulary speech recognition,” *Computer Speech & Language*, pp. 1237-1265, 1990.

- [24] Wilpon J., L. Rabiner, C. Lee, E. Goldman, “Automatic Recognition of Keywords in Unconstrained Speech Using Hidden Markov Models,” *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. 38, no. 11, 1990.
- [25] Ferguson J., “Hidden Markov Analysis: An Introduction in Hidden Markov Models for Speech,” Institute for Defense Analyses, Princeton, 1980.
- [26] Levinson S., L. Rabiner, M. Sondhi, “An Introduction to the Application of the Theory of Probabilistic Functions of a Markov Process to Automatic Speech Recognition,” *Bell Syst. Tech. J.*, vol. 62, no. 4, pp. 1035-1074, 1983.
- [27] Rabiner L., B. Juang, “Statistical Methods for the Recognition and Understanding of Speech,” *Encyclopedia of Language and Linguistics*, 2004.
- [28] Baum L., “An Inequality and Associated Maximization Technique in Statistical Estimation for Probabilistic Functions of Markov Processes,” *Inequalities*, vol. 3, pp. 1-8, 1972.
- [29] Theodoridis S., K. Koutroumbas, “Pattern Recognition: Second Edition,” Elsevier Academic Press, 2003.
- [30] Baum L., “An Inequality and Associated Maximization Technique in Statistical Estimation for Probabilistic Functions of Markov Processes,” *Inequalities*, vol. 3, pp. 1-8, 1972.
- [31] Poritz A., “Linear Predictive Hidden Markov Models and the Speech Signal,” in Proc. ICASSP-82, Paris, 1982.
- [32] Liporace L., “Maximum Likelihood Estimation for Multivariate Observations of Markov Sources,” *IEEE Trans. on Information Theory*, vol. 28, no. 5, pp. 729-734, 1982.
- [33] Juang B., S. Levinson, M. Sondhi, “Maximum Likelihood Estimation for Multivariate Mixture Observations of Markov Chains,” *IEEE Trans. Information Theory*, vol. 32, no. 2, pp. 307-309, 1986.
- [34] Juang B., “Maximum Likelihood Estimation for Mixture Multivariate Stochastic Observations of Markov Chains,” *AT&T Tech. J.*, vol. 64, no. 6, pp. 1235-1249, 1985.
- [35] Mohri M., “Finite-State Transducers in Language and Speech Processing,” *Computational Linguistics*, vol. 23, no. 2, pp. 269-312, 1992.
- [36] McCullough W., W. Pitts, “A Logical Calculus of Ideas Immanent in Nervous Activity,” *Bull. Math Biophysics*, vol. 5, pp. 115-133, 1943.
- [37] Lippmann R., Review of Neural Networks for Speech Recognition, Readings in Speech Recognition, 1990.
- [38] Juang B., C. Lee, W. Chou, “Minimum Classification Error Rate Methods for Speech Recognition,” *IEEE Trans. Speech & Audio Processing, T-SA*, vol. 5, no. 3, pp. 257-265, 1997.
- [39] Vapnik V., Statistical Learning Theory, John Wiley and Sons, 1998.
- [40] Lee K., Large-vocabulary Speaker-independent Continuous Speech Recognition: The Sphinx System, Ph.D. Thesis, Carnegie Mellon University, 1988.
- [41] Schwartz R., C. Barry, Y. Chow, etc., „The BBN BYBLOS Continuous Speech Recognition System,” in Proc. of the Speech and Natural Language Workshop, Philadelphia, 1989.
- [42] Murveit H., M. Cohen, P. Price, etc., „SRI's DECIPHER System,” in proceedings of the Speech and Natural Language Workshop, 1989, Philadelphia.
- [43] Young S., „the HTKBook,” <http://htk.eng.cam.ac.uk/>.

- [44] Glass J., E. Weinstein, „SpeechBuilder: Facilitating Spoken Dialogue System Development,” 7th European Conf. on Speech Communication and Technology, Aalborg Denmark, 2001.
- [45] Zue V., “Jupiter: A Telephone-Based Conversational Interface for Weather Information,” *IEEE Trans. On Speech and Audio Processing*, vol. X, pp. 100-112, 2000.
- [46] Gorin A., B. Parker, R. Sachs, J. Wilpon, “How May I Help You?,” 1996.
- [47] Kishorjit N., R. Vidya, Y. Nirmal, B. Sivaji, “Manipuri Morpheme Identification,” in *Proceedings of the 3rd Workshop on South and Southeast Asian Natural Language Processing*, Mumbai, 2012.
- [48] Huang X., A. Acero, H. Hon, Spoken Language processing – A Guide to Theory, Algorithms and System Development, Prentice Hall PTR, 2001, pp. 375-407.
- [49] Потемкин В., В. Медведев, Нейронные сети. MATLAB 6, Диалог-МИФИ, 2002.
- [50] Младенов В., С. Йорданова, Размито управление и невронни мрежи, София, 2006.
- [51] Тренчев И., П. Миланов, Н. Пенчева, И. Мирчев, Невронни мрежи, Благоевград: ЮЗУ "Неофит Рилски", 2010.
- [52] Георгиева П., Генетични размити ситеми, Бургас: Полиграф, 2016.

FRACTAL MODELS IN BIOLOGY

Nikolova Stefani, Burgas Free University, stephany_951@mail.bg

Georgieva Penka V., Burgas Free University, pgeorg@bfu.bg

Abstract: Along with the progress in biological structures modeling and simulating their development, the visualization of simulation results develops and this leads to a better understanding of morphogenesis and launches new procedural techniques for realistic image synthesis. The application of fractals is extremely useful both for the measurement of the dimensional properties and spatial parameters of the unregularly biological structures, as well as for understanding of the structural/morphological organization of the living tissues and organs.

This article proposes a fractal model of human lung and an algorithm for 3D visualization of this model is presented.

Keywords: fractal, fractal modeling, lung modeling, 3D modeling, bifurcation design

ФРАКТАЛНИ МОДЕЛИ В БИОЛОГИЯТА

Стефани Николова, Бургаски свободен университет, stephany_951@mail.bg

Пенка В. Георгиева, Бургаски свободен университет, pgeorg@bfu.bg

Абстракт: Заедно с напредъка в моделирането на биологически структури и симулацията на тяхното развитие се развива визуализацията на стимулационните резултати, което води до по-доброто разбиране на морфогенезата и поставя началото на нови процедурни техники за реалистичен синтез на изображението. Прилагането на фракталните закономерности е изключително полезно както при измерването на дименсионните свойства и пространствените параметри на неправилните биологични структури, така и за разбирането на структурната/морфологична организация на живите тъкани и органи.

В тази статия е предложен фрактален модел на човешки бял дроб и е представен алгоритъм за алгоритъм за 3D визуализацията на този модел.

Ключови думи: фрактал, фрактално моделиране, модел на бял дроб, 3D моделиране, бифуркационен дизайн

I. ВЪВЕДЕНИЕ

Терминът *фрактал*, въведен от Манделброт, произхожда от латинското прилагателно *fractus*=изменен, неправилен, чийто корен е латинския глагол *frangere*=да се начупи, да създава неправилни участъци. За да бъдат наречени фрактали, биологичните и/или природни обекти трябва да изпълняват определен брой теоретични и методологични критерии: високо ниво на организация, неправилност на формата, функционално морфологично самоподобие, инвариантност, итеративност. Откриването и изучаването

на фрактални закономерности е полезно при измерването на димензионните свойства и пространствените параметри на неправилни биологични структури, за разбиране на структурната/морфологична организация на живите тъкани и органи и за анализ след морфогенетични изменения. Фрактални форми се появяват при клетъчните мембрани, клетъчните ракови образувания, мозъчните тъкани и други [1]. Например свързването на вирус и жива клетка може да се опише с фрактали. Фракталната геометрия подпомага откриването на ракови клетки и изследването на инкубационния период на вируса на СПИН. Фракталният начин на растеж се използва при моделирането на някои морски организми (пр. корали). Мащабируемостта също е обект на интерес при изучаването на вегетативните екосистеми и поведението на гъсто населените популации.

Фракталната геометрия позволява разпознаването и количественото описание на сложни форми, живи организми, биологични тъкани и организирани структури от морфологични обекти, свързани чрез широката мрежа от функционални взаимодействия, които оформят адаптивността и правят възможен процеса на живот.

Биологичните елементи изразяват статистически самоподобни свойства и фрактални особености в рамките на дефиниран интервал от увеличения, наречен *прозорец на мащабиране*, където връзката между наблюдаваното увеличение и измерваната големина или дължина на обекта може да се установи и дефинира като фрактална размерност [2]. Фракталната размерност е статистическа мярка, свързана с морфологичната структурна сложност на клетъчните компоненти и биологични тъкани [3], като чрез фракталната размерност описва числено качествените морфологични черти и самоподобни характеристики на биологичните елементи.

Приложението на фракталната геометрия в клетъчната биология следва от откритието, че клетъчните мембранни системи притежават фрактални свойства, подобно на описания от Манделброт *Ефект на бреговата линия на Великобритания* [4]. Това дава отговор на въпроса, защо при увеличение на изображението измеренията на клетъчните мембрани дават по-високи стойности. [5]

Фракталният анализ е изключително полезен по отношение на електронната микроскопия за обективното изследване на крайни цитоплазмени структури и организацията на разнообразни видове хроматин, компоненти на клетъчното ядро, и други субклетъчни органели, както в нормални, така и при патологични тъкани и клетъчни образувания. При хистологията и цитологията, фракталната морфометрия, приложена при микроскопичното изследване на клетъчното ядро и ядрени компоненти, значително подобрява разбирането на клетъчното държание и диагнозата и прогнозата на разнообразни болестни състояния. [6], [7]

Количественото определяне на образуването на хроматина в ядрото на клетката от фракталната морфометрия се използва, да се определи степента на злокачественост при лечение на рака на гърдата [8]. Изследвания, насочени към периферията на клетъчните ядра, показват фрактални свойства, което прави възможно класификацията на ранните ракови образувания при яйчниците и да се различат нормалните от злокачествените клетки в черния дроб. [9]

II. ФРАКТАЛНА МЕТОДОЛОГИЯ

Според Манделброт *фракталното множество е множество от метричното пространство, за което размерността на Хаусдорф-Бесикович (D) е по-голяма от топологичната размерност (DT), т.е. $D > DT$.*

В природата фракталният обект се дефинира от своите структурни характеристики, а именно от вдлъбнатостта по повърхността и неправилността или отсъствието на

гладкост, инвариантност на формата или самоподобие, и сложност. Уравнението на Ричардсън-Манделброт осигурява математическа основа за разбирането на геометричните и фрактални структури, както и за тяхното измерване и интерпретиране:

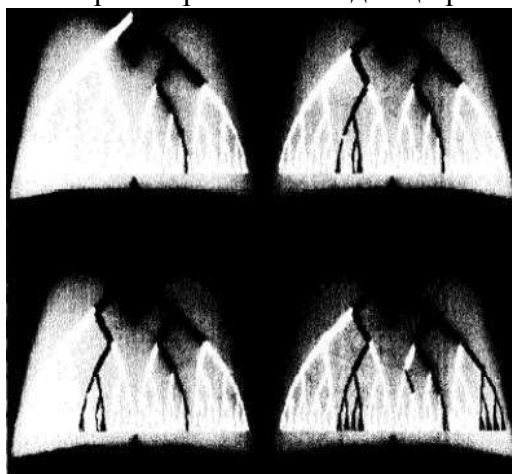
$$L(\varepsilon) = N(\varepsilon) \cdot \varepsilon, \quad (1)$$

където $L(\varepsilon)$ е дължината на контура на наблюдавания биологичен компонент, ε - мерната единица и $N(\varepsilon)$ брой мерни единици, нужни за покриването на контура. От $N(\varepsilon) = l_0^D \cdot \varepsilon^{-D}$, след заместване в (1), логаритмуване и преобразования следва равенството:

$$\ln \left[\frac{L(\varepsilon)}{l_0} \right] = (1 - D) \cdot \ln \left[\frac{\varepsilon}{l_0} \right], \quad (2)$$

което е уравнение на права с ъглов коефициент $1 - D$ в логаритмична скала. Дименсионната константа D може да бъде изчислена, откъдето се получава стойността на фракталната размерност.

Фракталните свойства на биологичните системи имат различна природа, произход и вид. В някои случаи геометричната форма на биологичния обект директно проявява фрактални характеристики, докато в други случаи тези свойства са скрити и могат да бъдат изучени, ако бъдат описани по специален начин [10]. За разлика от компактните обекти, фракталните имат много голямо повърхнинно лице. Белите дробове добре илюстрират тази особеност (фиг. 1). Разгърнатата площ на белия дроб е голяма колкото тенис корт, тъй като той е изграден от самоподобни разклонения по дължина, които образуват бронхиалното дърво [11]. Фракталната геометрия може да обясни промяната на диаметъра на бронхиалната тръба при всяко следващо разклонение на бронхите.



Фиг. 1. Снимки на бял дроб

Освен геометрията на бронхиалното дърво, с фракталната геометрия могат да бъдат описани и зависещите от времето особености на вдишването.

Кръвоносните съдове също имат фрактални свойства. Диаметрите на кръвоносните съдове (от капиляри до артерии) се изменят според разпределението на степенния закон, което е една от основните характеристики на фракталите.

Един от ярките примери за фрактален обект е повърхността на карфиола, където всяка една малка главичка е почти идентично копие на формата на цялото растение.

При моделирането и визуализиране на биологични структури се използва тяхната *морфогенеза*, а основни използвани технологии са *симулацията*, *визуализацията* на

симулационните резултати и фотореалистичното представяне на моделите. Морфогенезата, или развитието на сложни форми и закономерности в живите организми, осигурява множество примери за проявление на формата. По тази причина, нейните модели често изразяват значителен контраст между простотата на формообразуващите правила за поведение на индивидуалните компоненти и сложността на получените еволюционни процеси, шаблони, и форми.

Симулацията играе важна роля в изучаването на морфогенезата. Нейната роля е описана през 1952 година от Тюринг в [12].

Визуализацията на симулационните резултати улеснява тяхната интерпретация и се използва като метод за оценяване на модели. Тъй като липсва общоприета мерна единица за това, какво прави две форми (пр. дърветата) да изглеждат подобни, моделите се изследват визуално, като биват сравнявани с реалността.

Фотореалистичното представяне на моделите спомага в тяхното сравнение с естествените структури, и прави моделите полезни за приложения, синтезиращи изображения, като компютърната анимация, релефния дизайн и дигиталното изкуство.

Изучаването на морфогенезата се извършва в две направления. Първият тип се изразява в разглеждането на формата като производна на растежа [13], и е основано на факта, че формата на организма произлиза от нейния темп на *растеж* в различни посоки; следователно темпът на растеж заслужава да бъде изучавано като необходима предпоставка за теоретичното изследване на формата. Второто направление обръща погледом внимание върху движението на субстанциите през тъканта и е започнато от Тюринг [12].

III. ВИДОВЕ МОДЕЛИ НА БИОЛОГИЧНИ СТРУКТУРИ

1. Пространствено-ориентирани модели

1.1. *Реакционно-дифузионни шаблонни модели.* Този вид модели са разработен от Тюринг, за да се обясни "разпадането на симетрията и хомогенността", което води до появата на модели в изначално хомогенна, непрекъсната среда [12]. Шаблоните са резултат от взаимодействието между два или повече морфогена, които се разсейват в средата и влизат в химическо взаимодействие помежду си. Математически, този процес се изразява чрез система частни диференциални уравнения. Равномерното разпределение на морфогените е неустойчиво; случайните колебания се усилват, давайки като резултат устойчива схема с високи и ниски насищания, които графически се представят с различна цвятова гама. Мейнхард [14] синтезира реалистични изображения на черупки (фиг. 2), използвайки моделът на пигментация, разработен заедно с Клиндер [15].

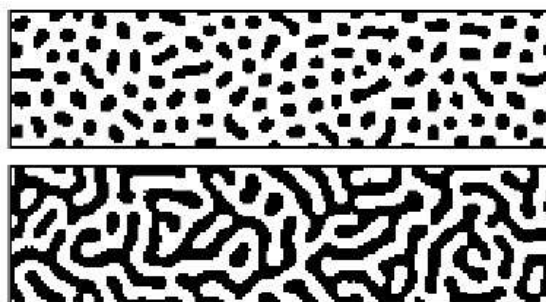


Фиг. 2. Фотография (ляво) и модел на Мейнхард (ясно) на *Natica enazona* (Fowler, Meinhardt, Prusinkiewicz, 1992)

1.2. *Дифузионно ограничен прираст на растежа.* При създаване на модела акцентът е върху постепенното разширяване на структурата и нейната граница, наречено прираст на растежа, като посоката на растеж се обуславя от мястото с по-голямо насищане на хранителни вещества. Кандорп прилага тримерен вариант на този процес, за да симулира и визуализира развитието на коралите и гъбите [16].

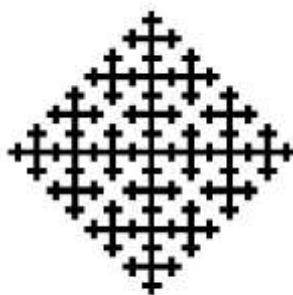
1.3. *Дифузионно ограничена агрегация.* Дифузионно ограничената агрегация има изразен фрактален характер на произлизащите разклонени структури.

1.4. *Клетъчна автомата.* Клетъчната автомата [17] е аналог на реакционнотифузионните модели в дискретно пространство. Йянг предлага клетъчно-автоматен модел на животинската шарка (фиг. 3), използвайки само две клетъчни състояния: с и без пигмент [18].



Фиг. 3. Животински марки, генерирани от Йянг
the reaction-diffusion model, proposed by Young

Улам прилага клетъчната автомата при симулация на развитие на дървовидните структури. На фигура 4 е изобразен модел, наречен *Малтийски кръстове*. При всяка итерация, моделът се разширява, освен ако между резултатните клонове няма да има колизия.

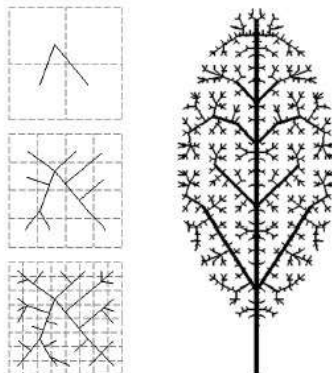


Фиг. 4. Дървовидна структура „Малтийски кръстове“
fractal structure generated by a cellular automaton operating on a square grid

1.5. *Вокселни автомати.* Тава е тримерен вариант на клетъчната автомата и се използва в компютърната графика при моделирането на аспекти от развитието на растенията, силно засегнати от околната среда. Впоследствие Грийн [19] разширява тази техника, за да улови вариациите в диаметъра на клоните и корените на дърветата. Създава симулационен модел на растежа на корените, които проправят своя път в почвата. В конкретния случай, вокселите съдържат информация за пътя на отделните елементи, които съставляват клонови и корени на дървото. Тази информация се използва, за да запази групи от нишки заедно и да направлява развитие им при появата на препятствия в околната среда.

1.6. *Развитие в разширяващото пространство.* Моделите, разгледани до тук могат да се развиват единствено по своя контур. Устойчивостта на основното пространство,

независимо дали то е непрекъснато или дискретно, предотвратява растежа във вътрешността. Готлиб [20] предлага геометричен модел на развитие, в който пространството се разширява равномерно (фиг. 5). Практическото приложение е растежа на листата, чиято съдова система се развива с цел поддържане на капацитета за преминаване на вода, хранителни вещества и продукти на фотосинтезата. Моделът има изразена йерархична организация, но все още има несъответствие между тяхното оформление и видовете, наблюдавани в природата. Точното моделиране на жилките по листата остава нерешен проблем.



Фиг. 5. Илюстрация на метода на Готлиб за генериране на шаблон
 Fig. 5. Illustration of the Gollub method for generating a template

2. Структурно-ориентирани модели

За разлика от пространствено-ориентирани модели, които описват цялото пространство, включително моделираната структура, структурно-ориентирани модели са съсредоточени върху развитието на отделни компонентите на структурата.

2.1 Системите на Линденмайер (*L-системи*) симулират развитието на линейни и мрежови структури, изградени от дискретни модули [21]. Параметричните *L-системи* позволяват изразяването на насищането на вещества, разпространяващи се в моделираната структура. Диференциалните *L-системи* разширяват *L-системите* до непрекъснатия времеви домейн, улесняващ компютърната анимация на процесите по развитие. Една *L-система* може да улови промените във формата в процеса на развитие.

2.2 *L-системи на съответствие* [22] са подобни на обикновените *L-системи*. Разликата се състои в граматиката, която използват и фактът, че съществува първоначална стартова повърхнина. Циклите на графите представляват клетъчни слоеве, чиято интерпретация е по-сложна от тази на дървовидните структури, тъй като наличието на цикли прави невъзможно прилагането на метрични свойства на модела, използвайки локални правила.

2.3 Флейшър и Бар [23] предлагат модел на *мобилни клетки в непрекъсната среда*, състоящ се от дискретни клетки, вградени в непрекъснат слой, който действа като среда, където химическите вещества се разпръскват, разсейват и влизат в реакция. Действията на клетките са разделени на непрекъснати процеси (растеж, движение) и дискретни събития (делене, създаване на дендрит, смърт). Клетките се движат под влиянието на физическите сили и взаимодействат с други клетки и слоеве чрез механични, химически и електрически средства. Вътрешно, дейността на всяка клетка се управлява от набор условни диференциални уравнения, зависещи от състоянието на клетката и локалната среда. Тези уравнения представляват "генетичната информация" на клетката и описват промените в масива от променливи, контролиращи клетъчното поведение.

IV. ФРАКТАЛНОСТ НА ЧОВЕШКИЯ БЯЛ ДРОБ

Белият дроб при бозайниците е разклоняваща се структура [24]. Въпреки, че скалируемите особености са били забелязани, Манделброт е първият, който изказва идеята, че структурата на белите дробове е фрактална. Впоследствие, фракталните концепции биват прилагани върху различни анатомични структури, включително разклонението на кръвоносните съдове, невронните мрежи, жлъчния и уринарния тракт, гънките на мозъка и тези на червата, сърдечните проводящи влакна, както и въздушните пътища.

Въздушните пътища в белия дроб се описват като дихотомно разклонено дърво, където родителската тръба се разделя в две дъщерни тръби от трахеята до крайните бронхиоли [25]. Всяка разклонение представлява ново поколение, като трахеята е нулево поколение, лавата и дясна тръба са първо поколение и т.н. Дихотомното разклоняване означава, че броят сегменти (N) във всяко следващо поколение (z) се е степен на двойката $N = 2^z$.

Класическите модели на мащабирането на въздушните пътища са експоненциални, взимайки предвид единствен характеризиращ фактор на скалиране. С други думи, средният диаметър или дължина на всяко поколение е умалена версия на предходното поколение посредством константа на скалиране, така че:

$$S(z) = S(0) \cdot x^z, \quad (3)$$

където $x < 1$ е коефициент на скалируемост, z е поредният номер на разклонение, S е размер на z -ото поколение и $S(0)$ е размер на трахеята. Сложността на структурата на бронхиалното дърво изисква по-пълна характеристика в сравнение с тази на експоненциалния модел.

Бифуркационен дизайн

Разклоненията на белия дроб се получават чрез асиметрична бифуркация с дървовидната архитектура, а връзката между диаметъра на родителския клон d_0 и диаметрите на клоновете от следващото поколение d_1 и d_2 , като $d_0 > d_1 \geq d_2$ е:

$$d_0^x = d_1^x + d_2^x, \quad (4)$$

където x е степента на съответната точка на разклоняване.

Бифуркационната архитектура се свежда до определянето на три параметъра:

- $\beta = \frac{d_1^2 + d_2^2}{d_0^2}$ - отношение на лицата на сеченията;
- $\gamma = \frac{d_2}{d_1} \leq 1$ - коефициент на асиметрия;
- x - степента на съответната точка на разклоняване.

След прилагане на (4) и преобразования, параметърът β може да бъде записан и по следния начин:

$$\beta = \frac{1 + \gamma^2}{(1 + \gamma^x)^{2/x}}. \quad (5)$$

От равенство (5) се определя ключовата важност на стойността на параметъра x , така например:

- ако $x > 2$ и γ се увеличава, то β също се увеличава;
- ако $x = 2$, то β не зависи от γ ;
- ако $x < 2$ и γ се увеличава, то β намалява.

Фрактална размерност на базата на бифуркационния дизайн

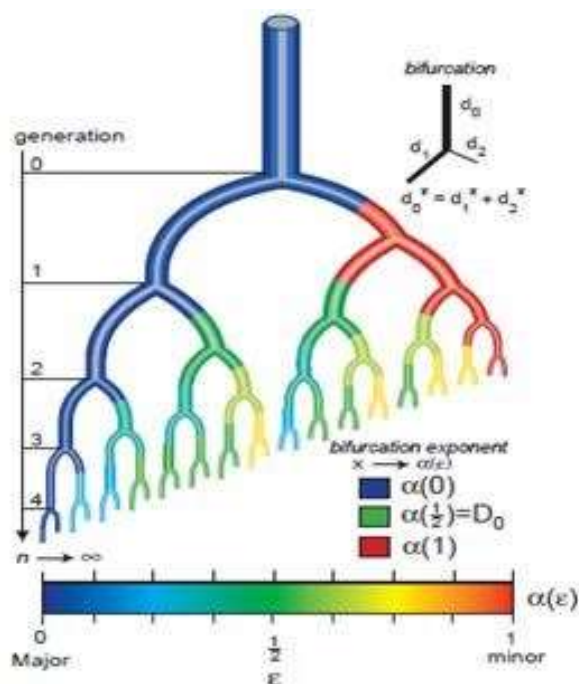
Бифуркационният дизайн се приема като формираща основа на фракталната размерност при дървовидните структури, в случай че броят на поколенията нараства до безкрайност. Асиметричното разклоняване при биологичните мрежи е аналогично на бинарен канторов разклоняващ се процес с начален диаметър d_0 (фиг. 6). При n -тото поколение броят на разклоненията е $N = 2^n$, от които $\binom{n}{k}$ на брой са разклонения с диаметър $R_d^{-n}(\gamma R_d^{-1})^{n-k}$ при $k = 0, 1, 2, \dots, n$, тъй като разклоненията наляво имат поголеми диаметри R_d^{-n} , докато тези на ляво са с по-малки диаметри γR_d^{-n} .

За формалната дефиниция на фракталната размерност, се въвежда мерна единица M_x :

$$M_x = \sum_{k=0}^n \binom{n}{k} R_d^{-kx} (\gamma R_d^{-1})^{(n-k)x} = (R_d^{-x} + (\gamma R_d^{-1})^x)^n. \quad (6)$$

Тъй като при $n \rightarrow \infty$, то диаметърът $(\gamma R_d^{-1})^n \rightarrow 0$, откъдето M_x може да остане крайна единствено тогава и само тогава, когато $x = D_0$, където D_0 удовлетворява следното равенство:

$$R_d^{-D_0} + (\gamma R_d^{-1})^{D_0} = 1. \quad (7)$$



Фиг. 6. Фрактална архитектура

V. РЕАЛИЗИРАНЕ НА МОДЕЛ НА БЯЛ ДРОБ

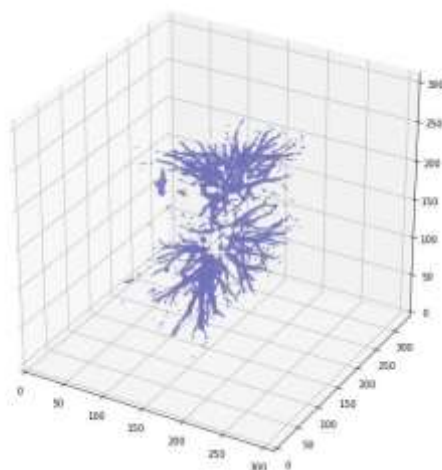
За да се пресъздаде реалистичен модел на бронхиалното дърво при човека първо се включват необходимите библиотеки и методи за визуализация и се извличат реални данни на пациенти (DICOM Library). Средата *IPython* се използва за реализиране на настоящия модел на въздушните пътища на белия дроб. За правилното сегментиране на желания участък от изображението се дефинира скалата на *Hounsfield unit* (HU) – линейна трансформация на първоначалното линейно измерване на коефициента на

отслабване в едно, при което радиоактивността на дестилирана вода при стандартно налягане и температура (STP) се дефинира като $HU = 0$, докато тази на въздуха е -1000 HU. При воксел със среден линеен коефициент μ , съответната HU стойност е както следва:

$$HU = 1000 \times \frac{\mu - \mu_{water}}{\mu_{water} - \mu_{air}},$$

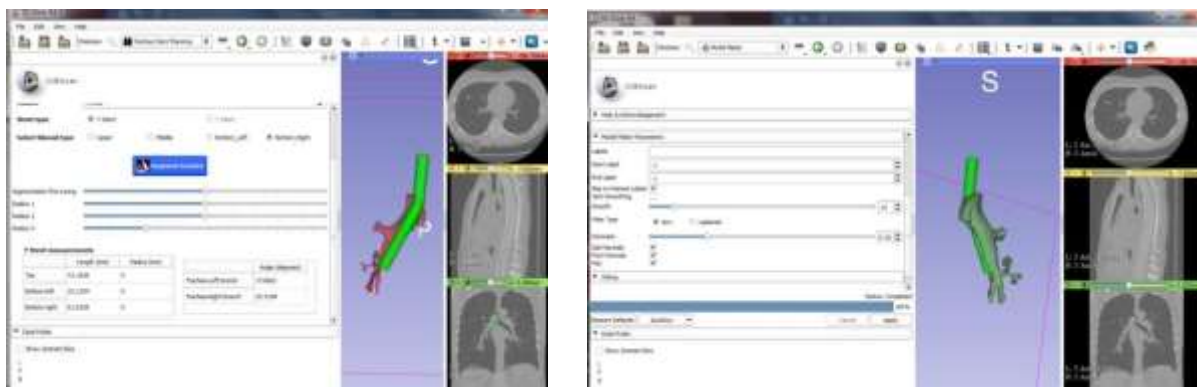
където μ_{water} и μ_{air} са съответните линейните коефициенти за водата и въздуха [26].

По-нататък се селектира пациент, за когото чрез функцията *plot_3d* се визуализират специфичните за него данни в тримерен вариант (фиг. 7).



Фиг. 7. Тримерна визуализация на дихателните пътища в средата IPython

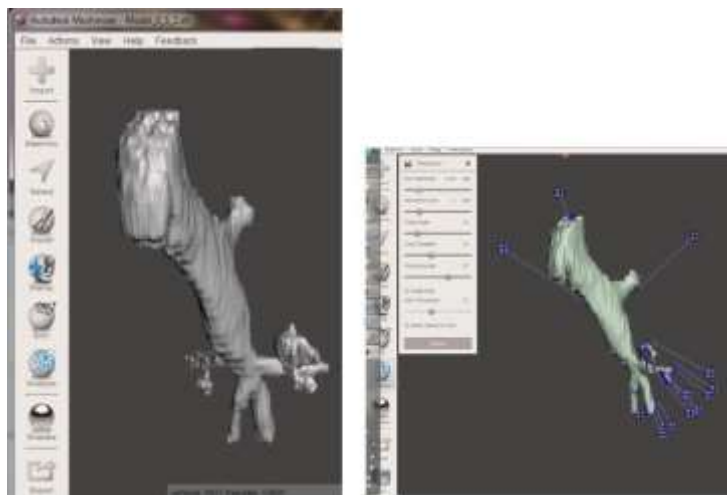
В следващия етап се използва *3D Slicer* за запазване на текущо състояние на отделните компоненти и се сегментират дихателните пътища чрез определяне на маркиращи точки (фиг. 8a) за избор на желания участък за сегментиране (фиг. 8b).



Фиг. 8. *3D Slicer* (a) Селектиране на маркиращи точки (b) Моделиране на дадения сегмент от дихателните пътища

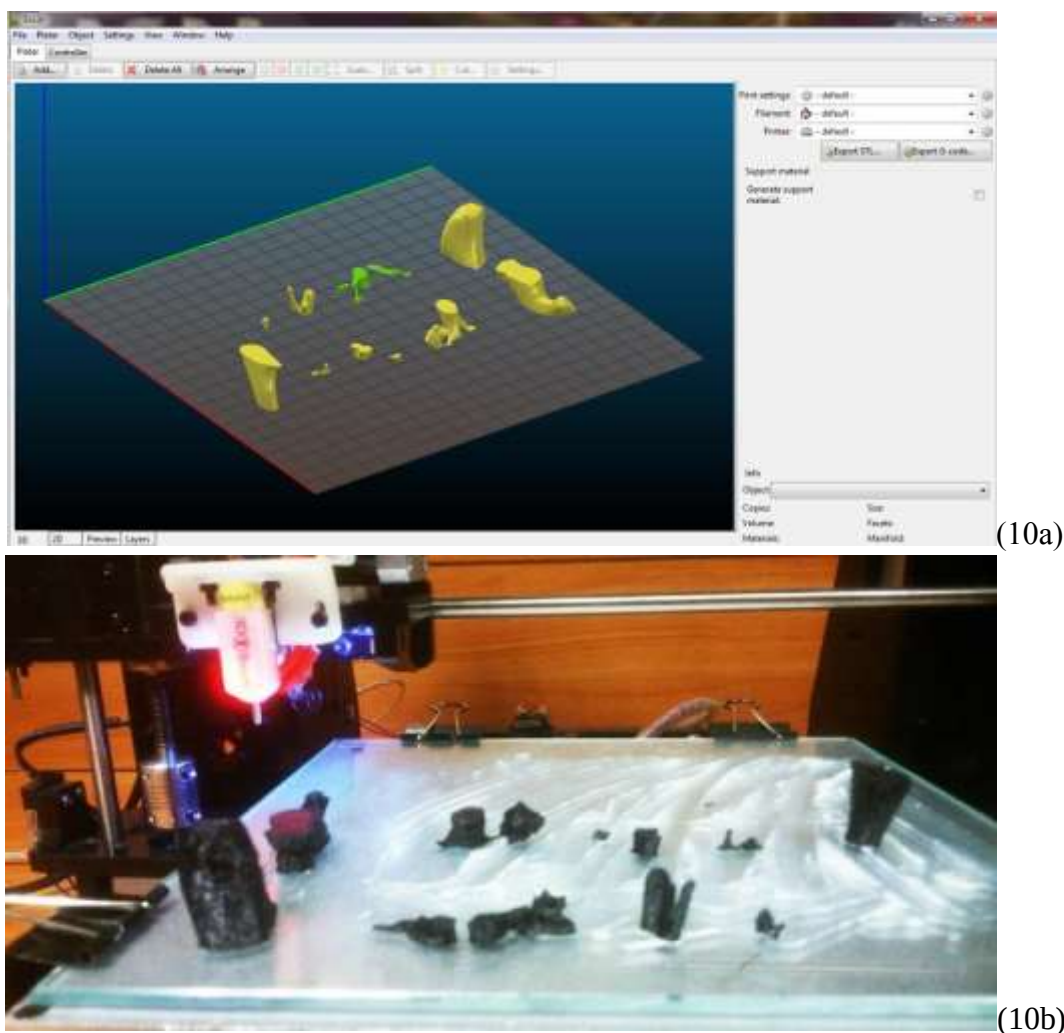
По-нататъшната обработка на файла се извършва в средата Autodesk Meshmixer. Интерфейсът е значително опростен, но предлага широка гама от възможности, някои от които ще бъдат разгледани в тази секция. Начинът, по който изглежда тримерният модел на селектирания участък от дихателните пътища се вижда на фигура 9a, а до него е показан първият анализ, на който обектът бива подложен (фиг. 9b).

Поради факта, че тримерният печат няма да стартира, ако моделът притежава некоригирани сегменти, то първо трябва да се премине през съответния анализатор на грешки, достъпен от главното меню *Analyze – Inspector*. В дадения случай, грешките са от тип липсваща повърхност. Тъй като несъответствията са повече на брой и засегнатите участъци са миниатюрни, за коригирането на всички тях се избира опцията *Auto Repair All*. Други възможности са: *Minimal Fill*, която ще запълни кухината с минимален брой полигони; *Flat Fill* – ще създаде плоска повърхнина; *Smooth Fill*, която ще слее всички заобикалящи повърхности.



Фиг. 9. Обработка на тримерен модел в *Meshmixer*. (a) Визуализация на обекта
(b) Анализ за грешки

Преди 3D принтирането, поради особеностите на принтера, готовият модел на бронхите трябва да бъде нарязан на по-малки части, притежаващи равна основа. Както се вижда на фигура 9, моделът не е поставен върху основа, а се намира свободно в пространството. Подобен тип модели биха могли да бъдат успешно отпечатани в първоначалния си образ чрез SLS технологията, но използваният принтер работи на принципа на FDM технологията, т.е. моделиране чрез отлагане на разтопен материал, то от тук идва необходимостта обектът да бъде предварително нарязан на по-малки сегменти (фиг. 10a). Това нарязва се извършва по оста Oz в отделна програма *Slic3r*. Готовият отпечатан модел се показан на фигура 10b.



Фиг. 10. Тримерен печат на сегмент от дихателните пътища.
(a) Подготвяне на модела в Slic3r (b) 3D принтиран резултат

VI. ЗАКЛЮЧЕНИЕ

Структурата на транспортните съдове в белия дроб проявява фрактална разклонена комплексност с произход, съответен на онези системи в природата, които притежават самоорганизирано критично състояние. Сложността на промените в белодробната циркулация е динамична, което предполага, че белите дробове са комплексна адаптивна система, в която липсва универсален фрактален дизайн и всъщност те не представляват строго самоподобна разклоняваща се мрежа. Също така има доказателства за фрактална самоафинна форма на мащабиране, включваща структурен организационен източник на податливост към съдови заболявания. Въпреки това, макар тази сложност да изглежда като причина за голяма част от разновидностите на дизайна на бифуркацията в белите дробове, тя не е отговорна за всички възможни модели, наблюдавани при разклоняването на морфогенезата и следователно, не трябва да се счита за единственият процес, управляващ биологичния разклонен модел на формообразуване.

Алтернативно, самоорганизираното критично състояние може да бъде част от универсалния модел на развитието на биологичната мрежа, растежа и адаптацията.

References

- [1] Losa A., “Fractals and their contribution to biology and medicine,” *Medicographia*, 34 (3), 2012.
- [2] Losa G., T. Nonnenmacher, “Self-similarity and fractal irregularity in pathologic tissues,” *Mod Pathol*, vol. 9, pp. 174-182, 1996.
- [3] Nonnenmacher T., G. Baumann, A. Barth, G. Losa, "Digital image analysis of selfsimilar cell profiles," *Int J Biomed Comput.*, vol. 37, pp. 131-138, 1994.
- [4] Mandelbrot B., “How long is the coast of Britain? Statistical self-similarity and fractional dimension,” *Science*, vol. 155, pp. 636-640, 1967.
- [5] Weibel E., "Design of Biological Organisms and Fractal Geometry," in *Fractals in Biology and Medicine*, vol. 1, Basel, Switzerland, Birkhäuser Press, 1994, pp. 68-85.
- [6] When B., W. Jacob, Van de Wouwer, et al., “Fractal dimension, form and shape factors for the quantification of nuclear signature profiles” *Fractals in Biology and Medicine*, 3, pp. 47-54, 2002.
- [7] Muniandy S., J. Stanlas, "Modelling of chromatin morphologies in breast cancer cells undergoing apoptosis using generalized Cauchy Field" *Comput Med Imaging Graph*, 32, pp. 631-637, 2008.
- [8] Einstein A., H. Wu, M. Sanchez, “Fractal characterization of chromatin appearance for diagnosis in breast cytology,” *J Pathol*, no. 185, pp. 366-381, 1998.
- [9] Nielsen B., F. Albregtsen, H. Danielsen, "Fractal signature vectors and lacunarity class distance matrices to extract new adaptive texture features from cell nuclei," *Fractals in Biology and Medicine*, vol. 3, pp. 55-65, 2002.
- [10] Halvin S., S. Buldyrev, A. Goldberger, etc., "Fractals in Biology and Medicine," *Chaos, Folutions & Fractals*, vol. 6, pp. 171-201, 1995.
- [11] West B., *Fractal Physiology and Chaos in Medicine*, Singapore: World Scientific, 1990.
- [12] Turing A., "The chemical basis of morphogenesis," *Philosophical Transactions of the Royal Society of London*, vol. 237, p. 37–72, 1952.
- [13] d’Arcy T., *On Growth and Form*, Cambridge: University Press, 1952. [14] Meinhardt H., *The Algorithmic Beauty of Sea Shells*, Springer, 2009.
- [15] Meinhardt H., M. Klinger, "A model for pattern formation on the shells of molluscs," *Journal of Theoretical Biology*, vol. 126, p. 63–89, 1987.
- [16] Kaandorp J., *Modeling growth forms of biological objects using fractals*, University of Amsterdam, 1992.
- [17] Toffoli T., N. Margolus, *Cellular Automata Machines: A new environment for modeling*, Cambridge: MIT Press, 1987.
- [18] Young A., "A local activator-inhibitor model of vertebrate skin patterns," *Math. Biosciences*, vol. 72, pp. 51-58, 1984.
- [19] Greene N., "Detailing tree skeletons with voxel automata," *SIGGRAPH’91*, 1991.
- [20] Gottlieb E., "The VT model: A deterministic model of angiogenesis," *IEEE Transactions on Biomedical Engineering*, 1993.
- [21] Lindenmayer A., "Mathematical models for cellular interaction in development," *Journal of Theoretical Biology*, vol. 18, p. 280–315, 1968.

- [22] Rozenberg G., A. Lindenmayer, "Parallel generation of maps: Developmental systems for cell layers," *Graph grammars and their application to computer science*, Springer, 1979, p. 301–316.
- [23] Fleischer K., A. Barr, "A simulation testbed for the study of multicellular development: Multiple mechanisms of morphogenesis," in *Artificial Life III*, Redwood City, AddisonWesley, 1993.
- [24] Mandelbrot B., *The Fractal Geometry of Nature*, W. H. Freeman and Company, 1982.
- [25] Rohler R., "Die Abbildungseigenschaften der Augenmedien," *Vision Res.*, vol. 2, pp. 391-429, 1962.
- [26] Weerakkody Y., K. Greenway et al., "Hounsfield unit," Radiopedia, [Online]. Available: <https://radiopaedia.org/articles/hounsfield-unit>.

REAL-TIME FACE RECOGNITION IN VIDEO STREAMING

Ivanov Ivelin, Burgas Free University, ivelin8910@gmail.com

Vanteva Violeta, Burgas Free University, v_vanteva@abv.bg

Georgieva Penka V., Burgas Free University, pgeorg@bfu.bg

Abstract: Face detection and recognition are areas in *Artificial Intelligence* that play a key role in information and people security and have various fields of application - control of groups of people, traffic monitoring, finding missing persons, searching for criminals and others.

In this study, a system for real-time face detection and recognition in video streaming is proposed, the system's model and implementation are described and test results are analyzed.

Keywords: face detection, face recognition, artificial intelligence, real-time software system

РАЗПОЗНАВАНЕ НА ЛИЦА ОТ ВИДЕО ПОТОК В РЕАЛНО ВРЕМЕ

Ивелин Иванов, Бургаски свободен университет, ivelin8910@gmail.com

Виолета Вантева, Бургаски свободен университет, v_vanteva@abv.bg

Пенка В. Георгиева, Бургаски свободен университет, pgeorg@bfu.bg

Абстракт: Засичането и разпознаването на лица са направления в *Изкуствения интелект*, които играят ключова роля за сигурността на информация и хора с разнообразни области на приложение - контрол на групи хора, наблюдение на трафик, откриване на изчезнали хора, издирване на криминално проявени лица и други.

В тази студия е предложена система за засичане и разпознаване на лица от видео поток в реално време, като са описани моделът и реализацията на системата и са анализирани резултати от проведените тестове.

Ключови думи: засичане на лица, разпознаване на лица, изкуствен интелект, софтуерни системи в реално време.

I. ВЪВЕДЕНИЕ

Разпознаването на лица е задача за откриване на шаблони в изображения, които са получени от различни източници на информация и по тази причина имат различни характеристики.

Ако изображението се променя във времето, то входните данн представляват видео поток.

Въпреки, че човешкото лице се състои само от две кости от две кости (череп, челюст) и 44 мускула, са възможни хиляди различни изражения на лицето. Умението за разпознаване на лица е централно и водещо за функционирането на всеки индивид, защото заема най-голям дял от посветената на обработването на зрителна информация функция (около 50% от мозъчната дейност).

Съществуват различни техники за идентифициране и удостоверяване, но основен аспект на разпознаването на лица е, че то е пасивна ненаатрапчива система за проверка и идентифициране на хора [1].

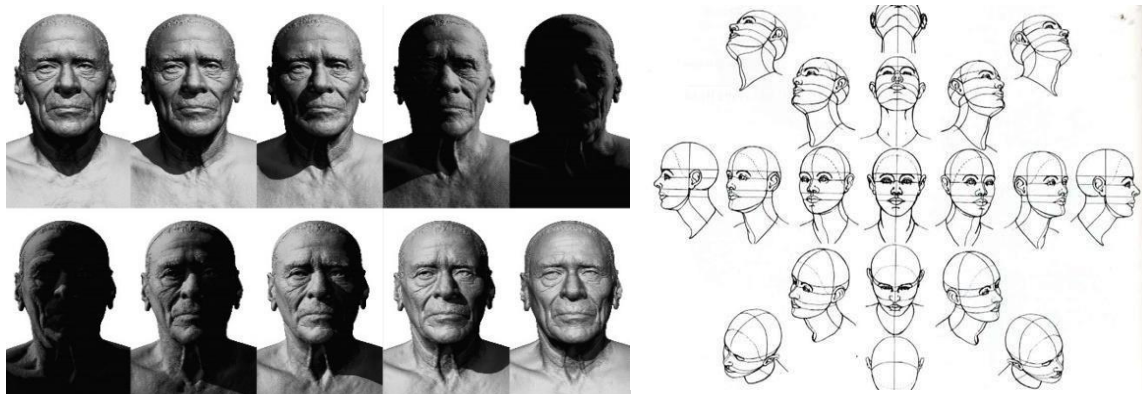
Автоматизираното разпознаване на лица започва развитието си през 60-те години на XX век. Първите полуавтоматични системи за разпознаване на лица са изисквали от администратора да открива различни отличителни белези като очи, уши, нос и уста, да изчисли разстоянията и съотношенията относно една референтна точка, след което резултатите са сравнявани с референтните налични данни [2]. През 1971 г. Голдщайн, Хармън и Леск използват 21 специфични субективни маркера (цвят на коса, дебелина на устни и др.), за автоматизиране на разпознаването [3]. Проблем на първите системи е, че както измерванията така и изчисленията са осъществявани ръчно. През 1987 г. Къбри и Сирович прилагат анализ на компонентите, което се счита за повратен момент, тъй като доказва, че са необходими по-малко от сто маркери за точно изчисление на подходящо подравнено и нормализирано изображение на лице [4]. През 1991 г. Пентланд и Търк откриват, че използването на собствени вектори може да даде възможност за откриване на лица в изображения и така да бъдат създадени надеждни автоматизирани системи за разпознаване на лица в реално време [5].

Днес технологиите за разпознаване на лица се използват за борба с паспортни измами, подкрепа на правосъдието, идентифициране на изчезнали хора, минимизиране на измамите, кражба на самоличност и други. [6]

Стандарти за изображения на лица

Съществуващите стандарти се отнасят предимно до съдържанието на изображенията и тяхното качество. За подобряване ефективността на системите за разпознаване е необходимо да се договорят стандарти, свързани с методологията на разпознаването [7]. Международният комитет за стандарти за информационните технологии (*International Committee for Information Technology Standards= INCITS*), който е признат от Организацията за развитие на стандарти за информационни технологии в САЩ (*Standards Development Organization=ANSI STO*) и Международната организация по стандартизация (*International Standards Organization=ISO*) чрез своята Международна електротехническа комисия (*International Electrotechnical Commission=IEC*) и Съвместния технически комитет (*Joint Technical Committee=JTC*) са издали 37 проекта за стандарти на биометричните данни. Международната организация за гражданска авиация е публикувала два основни документа, засягащи биометричното разгръщане на машинно четене на пътнически документи - ISO/IEC 19794-5-FDIS JTC 1/SC 37, M1/04-0041 (ANSI) INCITS (385). Стандартът M1 INCITS I2 приема или пълен или каноничен образ. Изображението на лицето трябва да включва цялата глава, косата, врата и раменете и ако се съхраняват и точки на характерните черти, то те базирани на SC29/MPEG4. Съществуват специфични изисквания за позата на лицето, отнасящи се до ъгъла на завъртане, наклон и посока на завъртане, изражението, осветлението и сенките. [8] ***Предизвикателства при***

разпознаване на лица от статични изображения Съществуват разнообразни варианти на изображение на човешкото лице, тъй като то зависи от фактори като поза, осветеност и изражение, от размера на лещата на обектива, както и от времето на излагане и отклонението на камерата. На фигура 1 са показани два от основните проблеми: различното осветление (фиг. 1а) и различен наклон на лицето (фиг. 1б). От друга страна, понякога се появяват малки междуличностни вариации (пр. идентични близнаци).



Фигура 1. Съществуващи предизвикателства в областта:

а) различната осветеност на лицето;

б) различен наклон на лицето.

За много от съществуващите системи за разпознаване предизвикателствата са големи и решени само до определена степен. Например, един подход за първоначално филтриране е налагането на ограничения в процеса на придобиване на изображения. Но има и случаи, в които вариациите сред изображенията на едно и също лице са по-големи от вариациите между изображенията на две различни лица, и тогава трябва да бъдат придобити по-прецизни изображения.

Предизвикателства при разпознаване на лица от видео поток

Допълнителните предизвикателства при разпознаване на лица от видео поток са следните:

- 1) претрупан фон с много ненужни детайли;
- 2) обектът за разпознаване липсва за даден период от време;
- 3) движението на обекта може да не е успоредно на оптичната ос;
- 4) локализиран се множество обекти;
- 5) различни ориентации на лицето на обекта;
- 6) загуба на детайлите на лицето поради размазване, причинено от движение на обекта спрямо камерата;
- 7) загуби, причинени от компресия на видео потока;
- 8) възникване на оптично изкривяване при изместване на обекта от оптичната ос и/или силно намалено разстояние до камерата.

Видове разпознаването на лица

Според целите на разпознаването има три основни вида разпознаване на лица.

1. Проверка и/или удостоверяване е задача, изискваща действие на потребителя под формата на заявка за самоличност, като въпросът е *Аз ли съм този, който казвам, че съм?*, а търсенето от вида *1:1*. Тестът за проверка се извършва чрез разделяне на лицата на две групи:

- *клиенти* – хора, които се опитват да получат достъп, използвайки собствената си самоличност;
- *измамници* – хора, които се опитват да получат достъп, използвайки фалшива идентичност, т. е. идентичност, позната на системата, но която не е тяхна.

2. Идентификация и/или разпознаване е задача, която не изисква взаимодействие с потребителя с въпрос *Кой съм аз?*. В този случай търсенето е от вида *1:n*, а идентификационният тест се основава на предположението, че всички образи са на вече известни личности.

3. Наблюдение е задача, която е обобщение на идентификационната задача с включване на неизвестни хора чрез търсене *1:n*.

Модули на система за разпознаване

Системите за разпознаване на лица имат три концептуални модула (фиг. 2):

- 1) модул за засичане на лице, в който се осъществява локализация в изображението чрез шаблони;



Фигура 2. Основни модули на система за разпознаване

- 2) модул за предварителна обработка, съдържащ нормализация на лицето, подравняване, преобразуване, завъртане, мащабиране, корекция на светлината и други. В този модул се извличат характеристиките на лицето като компактен

набор от дискриминационни геометрични и/или фотометрични характеристики на лицето чрез методи като: анализ на главните компоненти (PCA) [9], линеен дискриминационен анализ на Фишер (FLDA), прогнози за запазване на местоположението (LLP), невронни мрежи и други;

- 3) модул за класификация след сравняване на характеристичния вектор за съвпадение с лицеви изображения, вече записани в база данни с алгоритми като : най-близкия съсед, изкуствени невронни мрежи и други.

Съществуващи решения за разпознаване на лица

През последните 10 години са създадени множество разнообразни софтуерни системи за засичане и разпознаване на лица с различни приложения, но малко от тях работят в реално време. Популярни такива системи са: *digikam*, *FaceVACSVideoScan*, *Adobe Photoshop Lightroom*, *Picasa*, *DeepFace*, *FBI Sinister Next Generation Identification*, *Examiner*, *MFLO*, *NeoFace® Watch*.

Системи, работещи в реално време

Софтуерните системи, работещи в реално време стават все по-важни за функционирането на съвременното общество. Такива системи се използват за управление на въздушното движение, за разпознаване на обекти, за управление на командите, мрежови мултимедийни системи и други. При системите в реално време точността на поведението на системата зависи не само от логическите резултати от изчисленията, но и от физическия момент, в който се получават тези резултати [10]. Компютърна система, работеща в реално време, трябва да реагира на изменения в средата в рамките на предварително дефинирани интервали от време. Моментът, в който се генерира резултат, се нарича краен срок. Специално внимание се отделя на системите в строго реално време и гъвкаво реално време. Пропуснатият краен срок в такива системи е катастрофален и може да доведе до значителни загуби [11].

II. МОДЕЛ НА ПРИЛОЖЕНИЕ ЗА ЗАСИЧАНЕ И РАЗПОЗНАВАНЕ НА ОБРАЗИ В РЕАЛНО ВРЕМЕ

Обработка на видео поток

В описаната в тази студия софтверна система за засичане и разпознаване на лица в реално време, първоначално видео материалът се разделя на кадри, като на всеки 200 милисекунди се зарежда нов кадър от съответния видео материал и той бива обработен. Такъв интервал на бързодействие е избран, за да е съвместим с времето, за което нервен импулс стига от кожата на човек до мозъка (100 милисекунди) и времето за мигане (100-400 милисекунди), според базата данни на Харвард за стойности на биологични процеси [12].

Клъстеризация

Клъстеризацията на процеса на обработка на видео кадрите, използвана при реализирането на предложението в тази студия модел, има за цел оптимизация с цел намаляване времето за обработка на кадри и съответно увеличаване бързодействието на цялата система за засичане и разпознаване на обекти. Клъстерната архитектура за поддържане на голям брой потребители балансира натоварването, като всяка заявка е насочена към конкретен възел. [13]

Логическо описание на приложението

Приложението за разпознаване на лица се състои от четири основни модула (фиг. 3):

- 1) предварителна обработка на постъпващия видео материал;
- 2) клъстеризация на програмата;
- 3) засичане на образ в отделните кадри;
- 4) разпознаване на съответното засечено лице, ако такова е намерено.



Фигура 3. Логически модел на приложението

1. Предварителна обработка на постъпващия видео материал

След стартиране на приложението се осъществява връзка с камерата на всеки 200 милисекунди за приемане на кадри.

Всеки постъпил кадър се преобразува от цветно в черно-бяло изображение.

След това преобразование, кадърът се предава на следващия модул.

2. Клъстеризация на програмата

Софтуерното приложение асинхронно подава изображение на един от възлите на всеки 200 милисекунди, като решението избор на възел се взема в реално време от самата програма.

Всеки възел обработват конкретното изображение за 50-200 милисекунди. Броят на възлите на клъстера се определя от натовареността на системата в даден момент.

Основно предимство на разпаралеляването на тази част от алгоритъма за разпознаване е, че системата има възможност да обработва кадри, едновременно постъпващи от голям брой различни машини или устройства. След изпращането на

кадър до съответен възел, управлението се поема от следващия модул на приложението.

3. Засичане на образ в отделните кадри

Компонентът „засичане на образ“ се реализира отделно на всеки възел.

Съществуват различни алгоритми за засичане на обекти в изображение, като методите за засичане на лица са основно два вида:

намиране на лица в изображения с контролиран фон;

намиране на лица в изображения без ограничения за средата. При първия подход се използват изображения с обикновен монохроматичен фон или такъв, който е предварително зададен. Това улеснява премахването на фона, като след това остават само границите на самото лице, ако такова присъства в изображението. Вторият подход при засичането на лица е засичане без наличие на ограничения в средата. Използват се алгоритми, базирани на геометрични модели, разстояние на Хаусдорф [14] и други.

В настоящата система за засичане и разпознаване на лица е реализиран алгоритъмът на Виола-Джоунс [15], който се състои от четири етапа:

1. избор на характеристики на Хаар;
2. създаване на интегрално изображение;
3. обучение с *Адабууст*;
4. създаване на ефективни каскадни класификатори.

Този метод за засичане на лица използва зависимости, присъстващи в човешките лица, които се представят чрез черни и бели правоъгълни области. Тези правоъгълни характеристики се изчисляват чрез интегрално изображение $ii(x, y)$, при което на мястото на пиксел с координати (x, y) се записва сумата от пикселите, които се намират над и вляво от него:

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y'), \quad (1)$$

където $i(x, y)$ е оригиналното изображение.

Интегралното изображение се пресмята с едно обхождане на оригиналното изображение със следната двойка повторения:

$$s(x, y) = s(x, y - 1) + i(x, y) \quad (2)$$

$$ii(x, y) = ii(x - 1, y) + s(x, y), \quad (3)$$

при $s(x, -1) = 0$ и $ii(-1, y) = 0$.

Алгоритъмът за обучение *Адабууст* повишава класификационната способност на прост алгоритъм за обучение [16]. Алгоритъмът за обучение на еднослойна невронна мрежа претърсва множество от неврони и открива този с най-голяма класификационна грешка. За подсилване на обучението му се използва функция $h(x, f, p, \theta)$, дефинирана по следния начин:

$$h(x, f, p, \theta) = \begin{cases} 1, & \text{при } p \cdot f(x) < p \cdot \theta \\ 0, & \text{навсякъде другаде} \end{cases} \quad (4)$$

където f е елемент, p - полярност и θ – праг.

В последния етап на алгоритъма Виола-Джоунс се изгражда каскада от класификатори за намаляване на времето и повишаване на ефективността при засичане. Първо се прилагат по-простите класификатори с цел отхвърляне на поголямата част от подпрозорците, след което се използват по-сложните класификатори. Каскадата се изгражда от обучаващи класификатори отново чрез обучение *Адабууст*. Общата форма на процеса на засичане е дърво за вземане на решения, което в случая се нарича „каскада“. Положителен резултат от първия класификатор задейства изпълнението на втори класификатор, който също бива коригиран, за да постигне по-високи нива на засичане. Положителният резултат от втория класификатор задейства трети такъв и така до края на каскадата. Негативният резултат във всеки момент води до незабавното отхвърляне на текущия, установен като негативен подпрозорец. Този процес е илюстриран на фигура 4.



Фигура 4. Принцип на пораждаване на класификатори и отхвърляне на подпрозорци

Използването на каскадна структурата дава възможност за отхвърляне на възможно най-много подпрозорци във възможно най-ранна фаза. Фалшиво-положителната величина F на каскадата се изчислява по формулата:

$$F = \prod_{i=1}^K f_i, \quad (5)$$

където K е броят на класификаторите и f_i е фалшиво-положителната величина на i -тия класификатор на примерите, които достигат до него. Коефициентът на откриване D е следният:

$$D = \prod_{i=1}^K d_i, \quad (6)$$

където K е броят на класификаторите и d_i е степенята на засичане на i -тия класификатор на примерите, които достигат до него.

Като се имат предвид конкретните цели за общите фалшиво-положителни стойности и коефициентите на засичане, целевите нива могат да бъдат определени за всеки етап

от каскадния процес. Например, коефициент на засичане 0,9 може да бъде постигнат чрез 10- етапен класификатор, ако всеки етап има степен на откриване от 0,99.

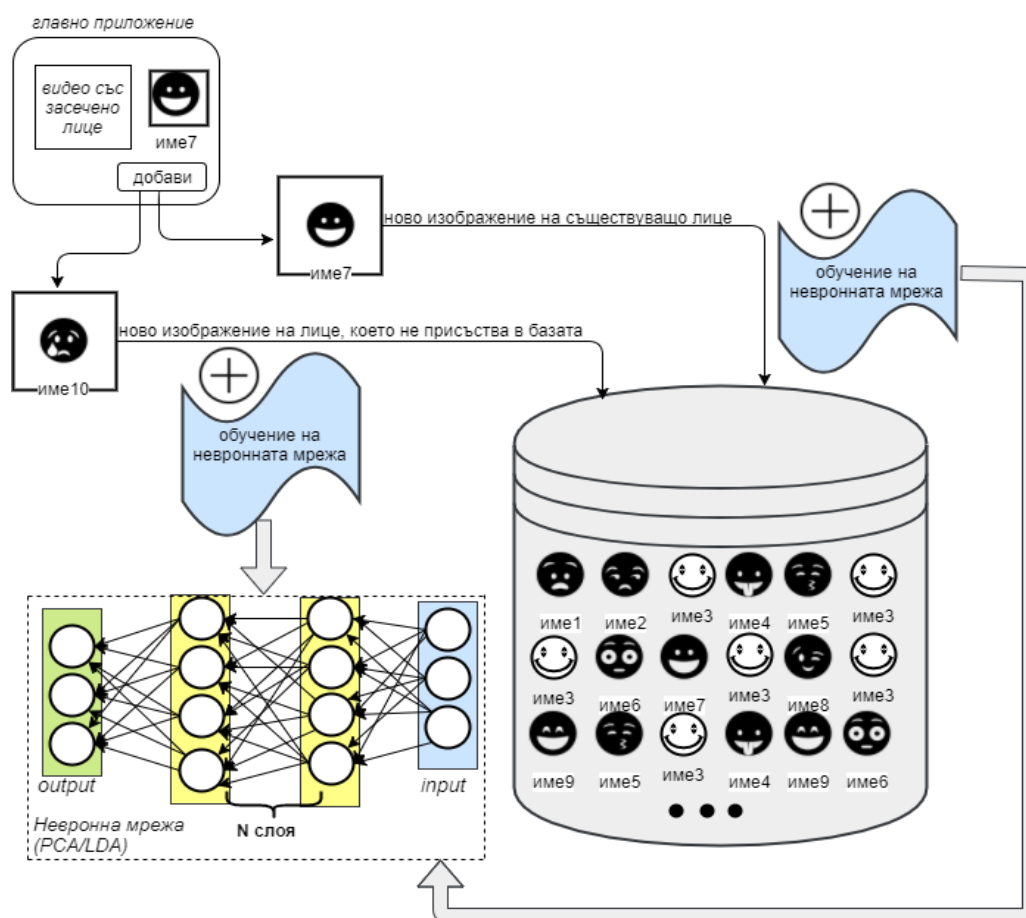
Броят на характеристиките, оценени при обработване на реални изображения е вероятностен процес. Очакваният брой характеристики N , които се оценяват е:

$$N = n_0 + \sum_{i=1}^K \left(n_i \prod_{j<i} p_j \right), \quad (7)$$

където K е броят на класификаторите, p_i е положителната скорост на i -тия класификатор и n_i е броят на характеристиките в i -тия класификатор.

4. Разпознаване

В този модул на приложението се постигат няколко основни задачи. Първо се съставя база от изображения в реално време, като след засичането на лице се предоставя възможност потребителят да въведе името му и това лице автоматично се добавя в базата от данни с тестови изображения (фиг. 5). След това невронната мрежа се обучава над новополучената обучителна база, като това се осъществява едновременно със засичането. Алгоритмите, които са използвани в невронната мрежа имат за основна цел да премахнат общите черти и да сравняват засеченото лице с уникалните черти с тези, които са вече разпознавани от системата. Това се постига с *Анализ на главните компоненти (АГК)* [17].



Фигура 5. Принцип на работа на модула за разпознаване

Анализ на главните компоненти (АГК)

Нека в базата от данни има n на брой обекта, като за всеки от тях се наблюдават m на брой характеристики. Тези наблюдения се записват в базата от данни като n на брой m -мерни вектори $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n$. По този начин всички наблюдения могат да се разглеждат като матрица X от тип $m \times n$. В АГК се използва следната теорема.

Теорема. (спектрална теорема) Ако за матрицата A от тип $n \times n$ е изпълнено $A = A^T$, то:

- 1) всички собствени стойности $\lambda_1, \dots, \lambda_n$ на матрицата A са реални числа и съществуват ненулеви собствени вектори $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ с реални координати, такива че $A \cdot \vec{v}_j = \lambda_j \cdot \vec{v}_j$, за всяко $j = 1, 2, \dots, n$;
- 2) собствените вектори, съответстващи на различни собствени стойности са ортогонални;
- 3) съществува диагонална матрица $D = \text{diag}(\lambda_1, \dots, \lambda_n)$ от тип $n \times n$ и ортогономална матрица U от тип $n \times n$ с елементи по стълбове съответните координати на собствените вектори, за които е изпълнено следното равенство:

$$A = U \cdot D \cdot U^T$$

В системата за разпознаване тази теорема се прилага върху симетричните матрици $X^T \cdot X$ от тип $n \times n$ и $X \cdot X^T$ от тип $m \times m$. Съществува връзка между собствените стойности и собствените вектори на $X^T \cdot X$ и $X \cdot X^T$. Наистина, нека $\vec{v}$ е ненулев собствен вектор на матрицата $X^T \cdot X$, съответстващ на собствената стойност $\lambda \neq 0$. Тогава е в сила равенството:

$$(X^T \cdot X) \cdot \vec{v} = \lambda \cdot \vec{v}, \quad (8)$$

откъдето след умножаване отляво с X и прилагане на свойствата за умножение на матрици се получава равенството:

$$X \cdot X^T \cdot (X \cdot \vec{v}) = \lambda \cdot (X \cdot \vec{v}) \quad (9)$$

От (9) следва, че $X \cdot \vec{v}$ е собствен вектор на матрицата $X \cdot X^T$, а λ е съответната собствена стойност. Доказателството за $X \cdot X^T$ е аналогично. [18]

Доказаното свойство е особено важно в случай, че Това предположение, което бе и доказано, е доста важно в случая, в който стойностите на m и n се различават значително (пр. $m = 500$ и $n = 2$).

Друго важно свойство, което се използва в системата за разпознаване е фактът, че собствените стойности на матриците $X^T \cdot X$ и $X \cdot X^T$ са неотрицателни числа. Наистина,

нека $\vec{v}$ е ненулев собствен вектор на матрицата $X^T \cdot X$, съответстващ на собствената стойност $\lambda \neq 0$. Тогава за дължината на вектора $X \cdot \vec{v}$ е в сила:

$$|X \cdot \vec{v}|^2 = (X \cdot \vec{v})^T \cdot (X \cdot \vec{v}) = \vec{v}^T \cdot (X^T \cdot X) \cdot \vec{v} = \vec{v}^T \cdot (\lambda \cdot \vec{v}) = \lambda \cdot (\vec{v}^T \cdot \vec{v}) = \lambda \cdot |\vec{v}|^2 \geq 0, \quad (10)$$

Следователно $\lambda \geq 0$.

Нека векторът $\vec{\mu}$ е средноаритметично на векторите $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n$:

$$\vec{\mu} = \frac{1}{n} (\vec{x}_1 + \vec{x}_2 + \dots + \vec{x}_n), \quad (11)$$

вариацията $\text{Var}(X)$ е:

$$\text{Var}(X) = \frac{1}{n-1} ((\vec{x}_1 - \vec{\mu})^2 + (\vec{x}_2 - \vec{\mu})^2 + \dots + (\vec{x}_n - \vec{\mu})^2) \quad (12)$$

и матрицата Y от тип $m \times n$ има за j -ти стълб вектора $\vec{x}_j - \vec{\mu}$ за $j = 1, 2, \dots, n$.

Конструира се матрица S от тип $m \times m$ по следния начин:

$$S = \frac{1}{n-1} Y \cdot Y^T, \quad (13)$$

за чиито елементи S_{kj} е изпълнено:

$$S_{kj} = \begin{cases} \text{вариацията на } k\text{-тата характеристика,} & \text{при } k = j \\ \text{ковариацията на } k\text{-тата и } j\text{-тата характеристики,} & \text{при } k \neq j. \end{cases}$$

От (13) следва, че матрицата S е симетрична и нека $\vec{u}_1, \vec{u}_2, \dots, \vec{u}_m$ са собствените ѝ вектори, отговарящи на собствените ѝ стойности, предварително подредени в низходящ ред $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$. Тези собствени вектори се наричат *главни компоненти* на X . От една страна следата T на матрицата S е равна на сумата вариациите, а от друга – на сумата от нейните собствени стойности ($T = \lambda_1 + \lambda_2 + \lambda_3 + \dots + \lambda_m$).

Тогава:

- посоката на първия главен компонент $\vec{u}_1$ отчита значимостта на λ_1 в общата вариация, равна на $\frac{\lambda_1}{T}$; посоката на втория главен компонент $\vec{u}_2$ отчита значимостта $\frac{\lambda_2}{T}$ на λ_2 и т.н.;
- от $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ следва, че векторът $\vec{u}_1$ посочва най-значимата компонента;
- от всички ортогонални на $\vec{u}_1$ направления, векторът $\vec{u}_2$ посочва найзначимата компонента;
- от всички ортогонални на $\vec{u}_1$ и $\vec{u}_2$ аправления, векторът $\vec{u}_3$ посочва найзначимата компонента и т.н.

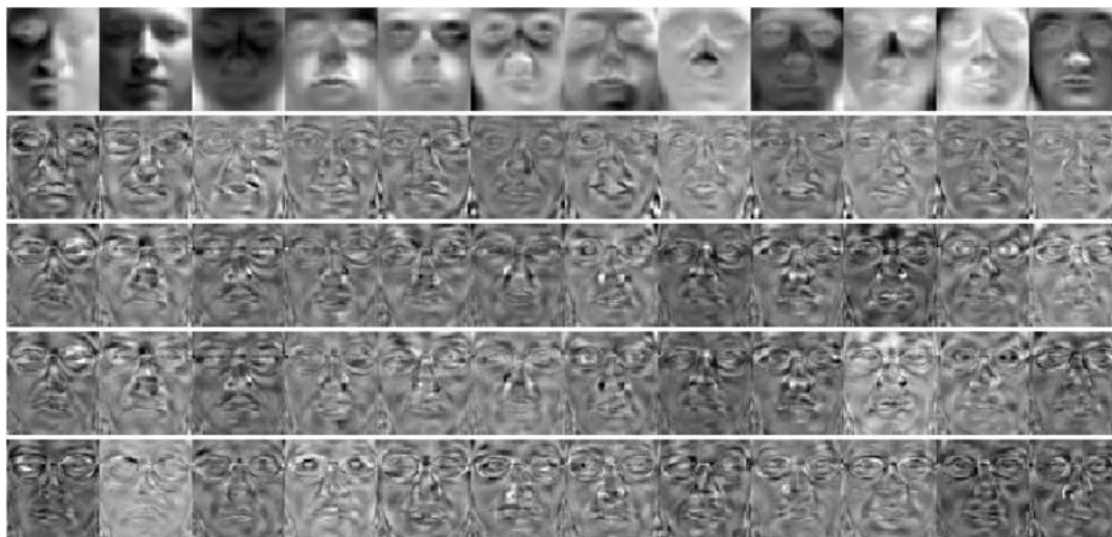
В приложенията, използващи *АГК*, първите собствените стойности на S се получават значително по-големи от всички останали. Така се оказва, че посоките примерно на първата и втората главна компонента обясняват *почти всички* вариации в данните.

Така, дори и да се отчитат много на брой характеристики, използването на две напълно достатъчно за получаване на състоятелни резултати.

Алгоритъмът на АГК се прилага в следните три основни стъпки:

1. събират се n на брой проби с m – мерни данни $\vec{x}_1, \vec{x}_2, \dots, \vec{x}_n$, изчислява се средната μ , конструират се матриците Y и S ,
2. пресмятат се собствените стойности на S , подредени в низходящ ред $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ и съответните собствени вектори $\vec{u}_1, \vec{u}_2, \dots, \vec{u}_m$,
3. анализират се получените резултати, като се отговаря на следните въпроси: дали малък брой от λ_i са много по-големи от всички останали, възможно ли е намаляване на размера на съществуващите данни, кои от променливите са най-важни и други.

Методът на главните компоненти е широко използван в разпознаването на лица, поради оптимизацията, която може да бъде постигната. На фигура 6 е показан примерен резултат, получен за обучителната база от изображения след прилагането на анализа на главните компоненти. За лицето на всеки човек от базата са останали само част от чертите, като това са само уникалните му характеристики. В случая първите главни компоненти са много по-значими от всички следващи.



Фигура 6. Обучителна база след прилагането на АГК

Невронната мрежа, използвана в конкретното работи идентично. След като невронната мрежа достигне до разпознаване, тя връща като изход име на разпознатото лице, ако е разпознато такова, заедно с вероятност, показваща степента на сигурност. Тази информация се изпраща пакетирани към главното приложение, което на базата на резултата, може да взема различни решения и също изписва в самия видео материал на екрана името на човека, който е разпознат, заедно с рамката, която е подадена като правоъгълник за мястото, на което е засечено лице.

III. РЕАЛИЗАЦИЯ НА МОДЕЛА

Използвани технологии за софтуерната реализация

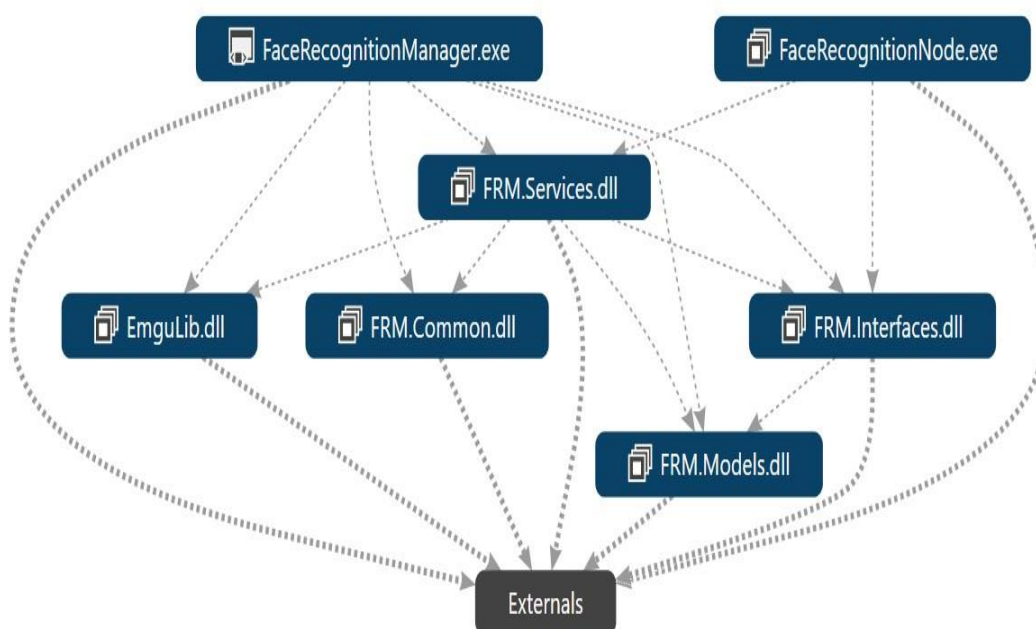
Разработката на система за разпознаване на лица включва софтуерно приложение, реализирано на езика за програмиране от високо ниво C# и платформата .NET.

Изборът на C# и .NET е обусловен от наличието на библиотеката с отворен код OpenCV (Open Source Computer Vision), подходяща за системи за засичане и разпознаване на лица. OpenCV е библиотека от функции за програмиране, насочени главно към компютърното възприятие в реално време с компоненти като машинно обучение, дървета на решения, спускане по градиента, алгоритъм k -ти най-близък съсед, поддържаща векторна машина (SVM), дълбоки невронни мрежи (DNN) и други.

Структура на приложението

Обща схема на приложението е показана на фигура 7. То се състои от два основни компонента:

- 1) **FaceRecognitionManager.exe** за вземане на кадри от източника тип „камера“ и изпращането им към подчинените му възли;
- 2) **FaceRecognitionNode.exe** за разпознаването на лице в съответния подаден кадър.



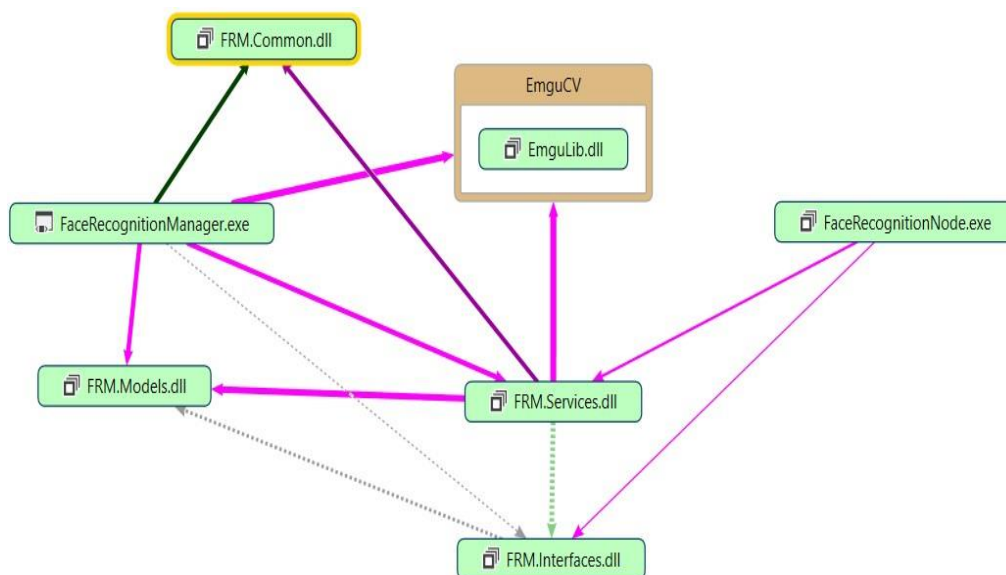
Фигура 7. Диаграма на връзките и проектите на системата за разпознаване

Архитектурата на приложението е многослойна:

- **FRM.Models.dll** съдържа всички нужни класове на приложението;

- **FRM.Common.dll** съдържа всички общи помощни функции;
- **FRM.EmguLib.dll** съдържа всички алгоритми, нужни за обработката и засичането на лице от снимка;
- **FRM.Interfaces.dll** съдържа всички интерфейси на класовете, като по този начин се осигурява независимост на главното приложение от конкретната имплементация и смяната ѝ, докато приложението е в работен режим „runtime“;
- **FRM.Services.dll** е библиотека, в която се съдържа цялостната логика нужна за създаване и поддържане на комуникацията между главното приложение и неговите възли.

EmguCV е Dll файл (динамична библиотека), представляващ интерпретация на библиотеката **OpenCV** в средата .NET на езика C#. Към **EmguCV** се обръщат приложенията **FaceRecognitionManager.exe** и **FRM.Services.dll** (фиг. 8).

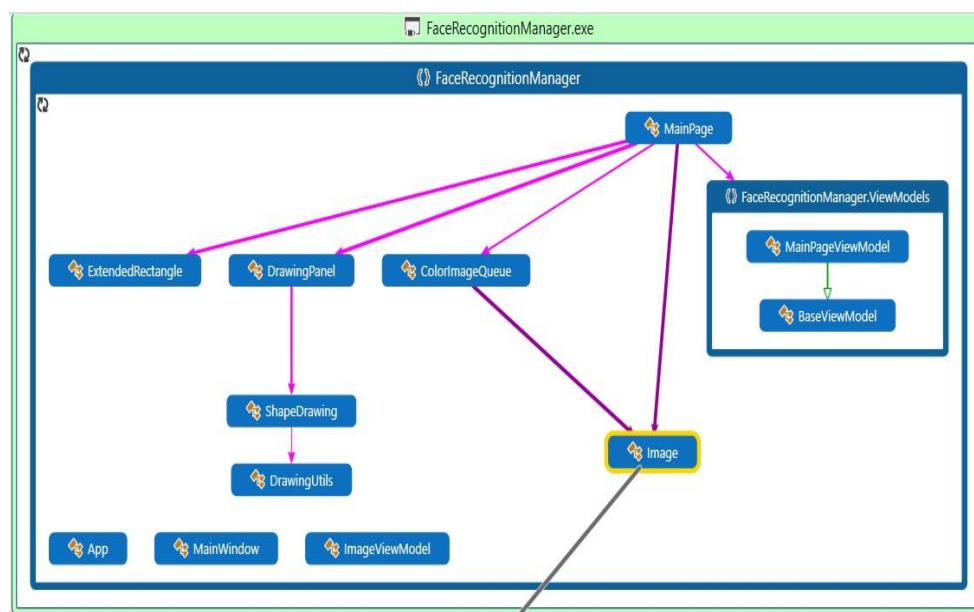


Фигура 8. Диаграма на приложенията в системата и връзките между тях и библиотеката с отворен код *EmguCV*

На фигура 9 е показана клас-диаграмата на главното приложение с име **FaceRecognitionManager**. Това приложение е изпълнима библиотека и се състои от няколко класа, служещи за визуализация резултатите на екрана:

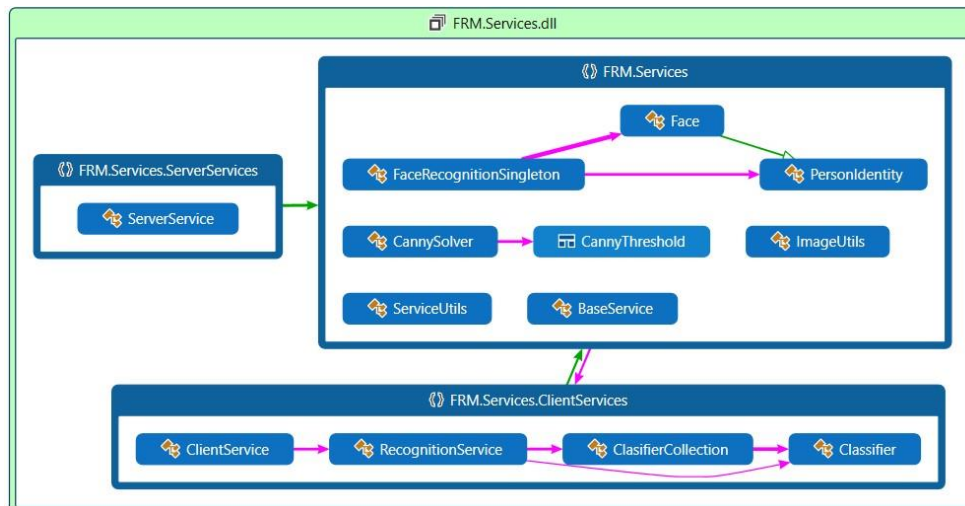
- **ExtendedRectangle** съдържа координатите на подаден правоъгълник, както и информация за типа му, определен от това дали разпознатият правоъгълник е на лице, око, уста и т.н.,
- Класовете **App**, **MainWindow** репрезентират съответното приложение и прозореца, който то създава, в който се добавят другите компоненти на интерфейса,

- **ShapeDrawing, DrawingUtils** са класове, съдържащи функции за рисуване на форми, текст и изображения на екрана,
- **DrawingPanel** – контролен панел, съдържащ нарисуваното изображение, като използва функционалностите предоставени от класа **ShapeDrawing**,
- **Image** – клас, съхраняващ модела на изображението в различните му варианти - цветно и черно-бяло, които се съхраняват и разглеждат като едно цяло,
- **ColorImageQueue** – клас за реализиране на циклична опашка с даден капацитет. Това нужно, поради това че процесът по вземане на кадри от устройството “камера” и процесът по изпращане на кадри за разпознаване са асинхронни. Така единият процес добавя елементи към опашката, а другият ги изважда и изпраща за разпознаване. В обработката на изображения в реално време се позволява даден кадър да се отхвърли, поради това че той вече е нерелевантен в даденият период от време т.е. не е нужно да обработваме стари кадри, които не отговарят на реалното състояние на средата в момента,
- **MainPage** – класът е репрезентация на страницата в приложението, в която са подредени всички визуални контроли. Реализиран е посредством MVVM модел, което е подход в софтуерните архитектури, осигуряващ двупосочна комуникация между данните и интерфейса. Така, ако някоя променлива си промени състоянието или стойността, това веднага се отразява на екрана, и обратно, ако състоянието на интерфейса се промени от взаимодействието на потребителя с него, то това ще промени стойностите на променливите в модела,
- **MainPageViewModel** съхранява всички данни, нужни за взаимодействието на потребителя с главната страница в приложението.



Фигура 9. Клас- диаграма на проекта FaceRecognitionManager, съдържащ главната функционалност

Следващият реализиран проект **FRM.Services** е главна част от реализираната система за разпознаване. Неговата структура е представена подробно на фигура 10. Тази библиотека се използва и в двете споменати вече приложения. Състои се главно от три namespace-a: един, използван в частта на сървъра/главния нод, един за частта на подчинените нодове/възли и един общ: *Клиентски клас, Сървърни класове*.



Фигура 10. Клас-диаграма на проекта Services

Реализираната многослойна архитектура на системата е необходима, за да се получи не само изрядност на кода, но и за да се реализират самостоятелно отделните логически модули.

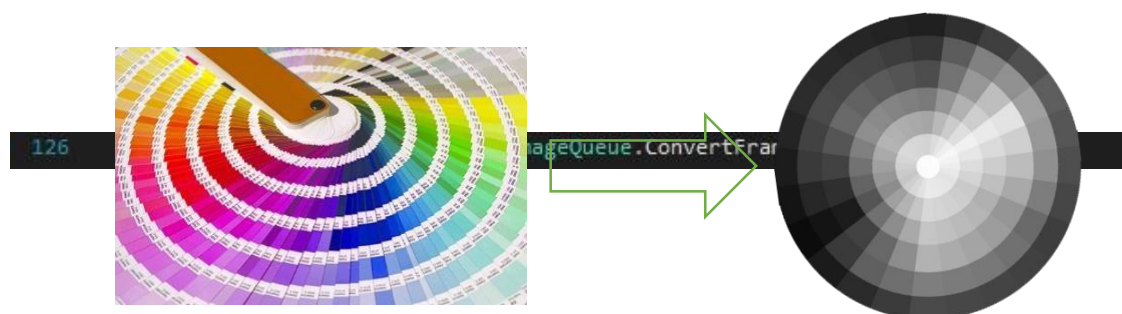
Осъществяване на комуникация с устройство за видео поток Първи необходим инструмент за да съществува системата за разпознаване, е източник на видео поток в реално време. Видео потокът може да бъде получен от камера на лаптоп или друга външна камера, като комуникацията със самото устройство се осъществява веднага при стартиране на приложението. Това е реализирано с DirectX пакета, който предоставя възможности за конструиране на графики на монитора и за получаване на изображения от външен източник (пр. камера), свързан с компютъра, както и за звук други. При обръщане към DirectX се създава канал за комуникация на камерата с приложението. Това се реализира чрез Windows драйверите на съответното устройство, като към тях се обръща именно DirectX. След осъществяване на въпросната комуникация, тя получава свой идентификационен номер в приложението, който се използва като обръщение към устройството (в случая камерата, вградена в компютъра). Чрез този идентификационен номер могат да се изпращат заявки към камерата в реално време от нея. След получаването на кадъра, той се визуализира в прозорец на приложението също в реално време. По този начин човешкото око възприема потока от кадри като непрекъснат видео стрийм в реално време. Програменият код, който реализира прихващането на кадъра от камера е:

```
122 1reference
123 private void Tick()
124 {
    var frame = capture.QueryFrame();
```

ICapture е интерфейс, който подпомага обръщението към камерата, като предоставя услуга за прихващане на кадър/изображение от видео поток.

Преобразуване на цветни изображения в черно-бели

Всеки кадър от видео потока се подлага на процедура по преобразуване на пикселите от цветната скала в черно-бяла (фиг. 11). На всеки пиксел от изображението съответства наредена тройка стойности, съответстващи на интензитета на червеното, зеленото и синьото от RGB скалата. Тези получени стойности се умножават със съответен коефициент, след което получените числа се сумират. От получената стойност се определя интензитета на сивото в интервала [0,255], като ако нула отговаря на черен цвят, а за 255 – на бял. От получените стойности се конструира RGB обект, който представлява полученият черно-бял цвят на съответния пиксел. По този начин се получава един цвят от трите интензитета и след като този алгоритъм се приложи върху цялото изображение, се получава чернобял кадър.



Фигура 11. Преобразуване от цветна в черно-бяла скала

Софтуерната реализация на тази функционалност се осъществява чрез извикване на функция, реализирана в класа **ColorImageQueue**. Програмният фрагмент за извикване на функция за цветово преобразуване е:

Преобразуването между двете скали се получа по формулата:

$$RedPixel * 0.3 + GreenPixel * 0.59 + BluePixel * 0.11,$$

Като полученият резултат се подава като параметър и по този начин се получава изображение, което има нюанси на сивото сивото.

Клъстеризация с цел оптимизиране

Клъстерът от възли в приложението е разработен с помощта на .NET технологията за уеб базирани функции WCF. Тази технология позволява напълно дуплексна асинхронна комуникация между главното приложение и отделните възли. Комуникацията се осъществява посредством TCP/IP протокол, който предоставя механизъм за сигурност и консистентно предаване на информацията, преминаваща през него. Протоколът предоставя и бързина спрямо алтернативните протоколи за комуникация.

Процес по засичане на лице в даден кадър

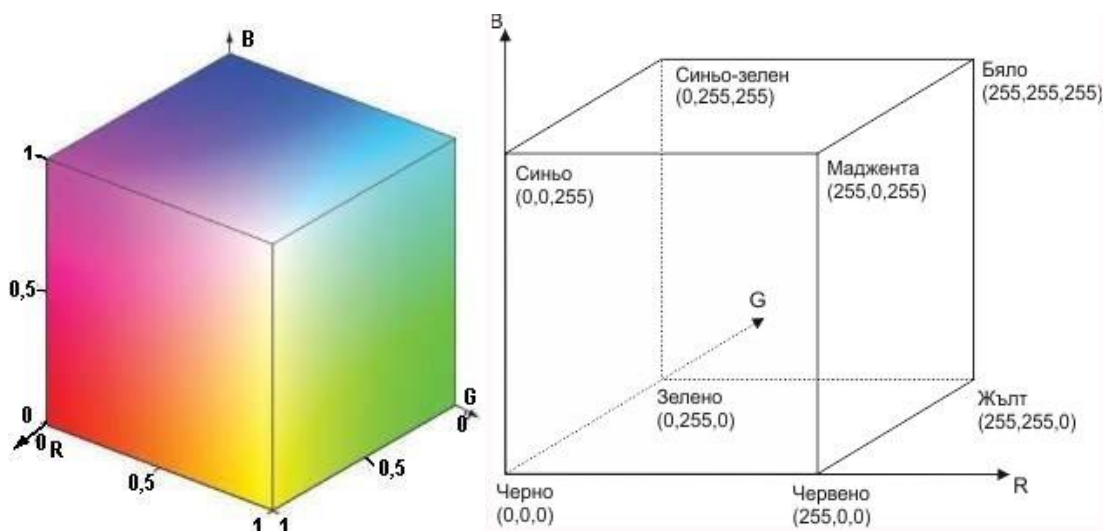
Процесът на засичането се състои главно в това да се изпрати извлечният и обработен предварително кадър на невронната мрежа, която реализира алгоритъма Виола-Джоунс. В приложението има различни опции за засичане. Те са свързани с това, че използваната библиотека предоставя намирането на различни характеристики на човешкото лице като функционалност. Включени са дескриптори:

- ✓ за засичане на очи,
- ✓ за наличие на очила,
- ✓ за засичане на ляво око,
- ✓ за засичане на дясно око,
- ✓ за засичане на лице анфас,
- ✓ за засичане на лице в профил,
- ✓ за засичане на цяло тяло,
- ✓ за засичане на долна част на тялото, ✓ за засичане на горна част на тяло, ✓ за засичане на усмивка.

Тези дескриптори са описани посредством XML файлове, които служат за йерархично структуриране на данни. Дескрипторът е описателен файл, който съдържа строго йерархично структурирани числа, които подпомагат действието на невронната мрежа. Преди самият взет вече кадър да се предаде на процеса за засичане, се прави проверка на това дали има клиенти. Ако все още има клиент, то кадърът се изпраща за засичане. Ако клиентът вече не използва в приложението, то текущият кадър се премахва от опашката с чакащи за проверка кадри, защото вече е ненужен.

Конструиране и поддържане на базата данни

Базата данни от изображения се съхранява във файлова система, като всички изображения автоматично се структурират по папки с имената на разпознатите хора за улесняване на повторно зареждане на изображенията и обучаване на алгоритъма за разпознаване на лица. Изображенията се пазят във формат PNG (Portable Network Graphics), с възможност за 16,7 милиона цвята на всеки пиксел (фиг. 12) и прозрачност, която се предоставя от четвърти слой.



Фигура 12. 3D RGB куб с цветови координати

Процес по разпознаване на засечено лице

Преди разпознаването на лицето, системата е осъществила засичането на лице и обучението на невронната мрежа, при наличието на ново лице в нея. Процесът по разпознаване се състои в избор на кой алгоритъм за разпознаване, като съответните данни се подават на невронната мрежа. Проверява се кой клиент е изпратил заявка за разпознаване на съответния кадър и дали има вече разпознато лице от всички засечени в кадъра. Накрая се изписва името на разпознатото лице на съответния кадър. Това име се запазва в променлива „name“, която след това се присвоява като характеристика на четириъгълник, съдържащ и координатите на намереното и разпознато лице. В края на разпознаването текущият четириъгълник се изважда от опашката с неразпознати все още, но засечени обекти.

За да се прекрати действието на системата към невронната мрежа, се извиква метод Predict, реализиран в класа-модел **FaceRecognitionSingleton**.

IV. РЕЗУЛТАТИ ОТ ИЗСЛЕДВАНЕТО

Използвани технологии за провеждане на тестове

Тестовите са направени на машина с процесор Intel i7 четвърта генерация 2.5GHz, като е симулирана връзка между много машини в клъстерната система. Параметрите на използваната камера са ключови за тестовите и резултатите, като всички тестове са проведени при следните характеристики:

- 60 градуса на полезрение;
- фиксиран фокус;
- 4мм фокално разстояние -;
- 30 кадъра/секунда при 640x480 резолюция.

В следващи тестове качеството на изображенията са намалени дори повече с цел да се покаже, че работоспособността на системата е в известна степен почти независима от параметрите на устройството.

Файлова база от данни

В първоначалната версия на приложението базата данни от изображения се описва с XML файлове, които представляват йерархична структура, която е лесна за четене от приложение, но все пак трудоемка за създаване. В текущата разработка е постигната елементарната структура на файловата база от данни. Изображенията се съхраняват в папки, като всяка папка се кръщава с името на човека, на когото принадлежат дадените снимки (фиг. 13).

Name	Date modified	Type
 Daniel	25-Nov-17 22:06	File folder
 Ivo	29-Oct-17 17:22	File folder
 Joro	25-Nov-17 22:06	File folder
 Mimi	29-Oct-17 17:23	File folder
 Stoian	25-Nov-17 22:06	File folder
 Valq	25-Nov-17 22:06	File folder
 Vili	03-Nov-17 16:44	File folder

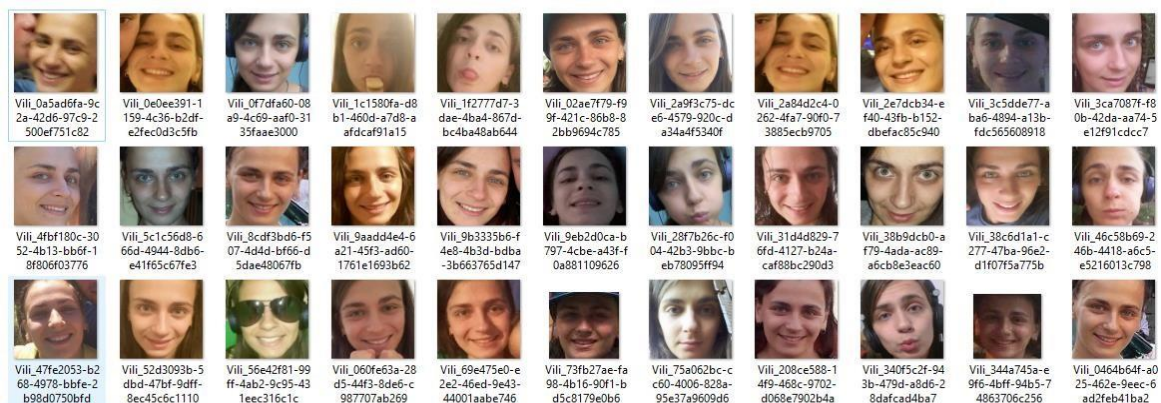
Фигура 13. Структура на файловата база данни

Цялата процедура по добавяне/премахване на папки и снимки е изцяло автоматизирана, т.е. се изпълнява от системата при необходимост.

В интерфейса на приложението са налични снимките, които се засичат от системата, заедно с поле за въвеждане на име на съответния засечен във видео потока човек. Това предоставя възможност за добавяне на нови изображения и хора в реално време, като допълнителни забавяния няма.

След всяко добавяне, невронната мрежа трябва да се обучи отново.

Приложението запазва вече изрязаните лица във формат .png с цел постигане максимално точно качество. Имената на тези изображения се конструират от системата, така че да бъдат използвани в процеса на разпознаване (фиг. 14).

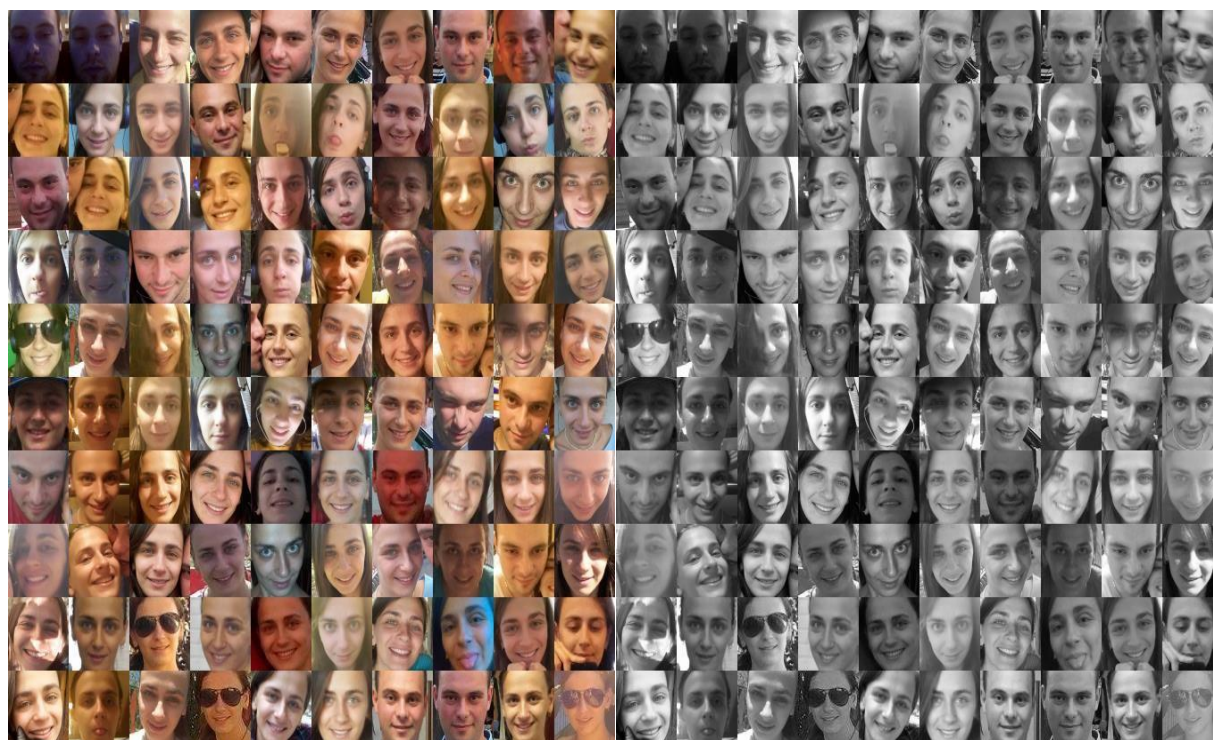


Фигура 14. Запазване на снимки със съответните имена

За ефективната работа на системата за разпознаване е необходимо да се събере възможно най-богата колекция от изображения, която да служат като база за тестовите. Важни е тези изображения да са с различни параметри и характеристики:

различен пол, различна форма на лицето, наличие на диоптрични очила, наличие на слънчеви непрозрачни очила, различни гримаси, наличие на шапка и различни прически на косата, различни ъгли на завъртане на главата.

Наличието на разнообразни изображения повишава възможностите на алгоритъма за разпознаване на лица, а една от съществените функционални възможности на предложената система е способността за засичане на лица с разнообразни характеристики. На фигури 15а) и 15б) са показани колажи от засечени във видео потока лица. При засичането изображенията се сканират и след това се изрязва и използва само частта, в която има лице. В паметта се пази снимката в цветен и черно-бял вариант, тъй като цветното изображение е нужно, за да се възстанови то при следващо показване на екрана в реално време, както и да присъства в базата данни, а черно-бялото е необходимо на алгоритмите както за засичане, така и за последващото разпознаване на лицето.



Фигура 15. Снимки на засечени от системата лица

а) цветни

б) черно-бели

На фигура 16 са показани други варианти на засечени от системата лица, като сред тях присъстват такива с диоптрични очила, усмихнати гримаси и други.



Фигура 16. Снимки на засечени лица с различни характеристики

Функционални възможности за засичане и разпознаване на множество лица

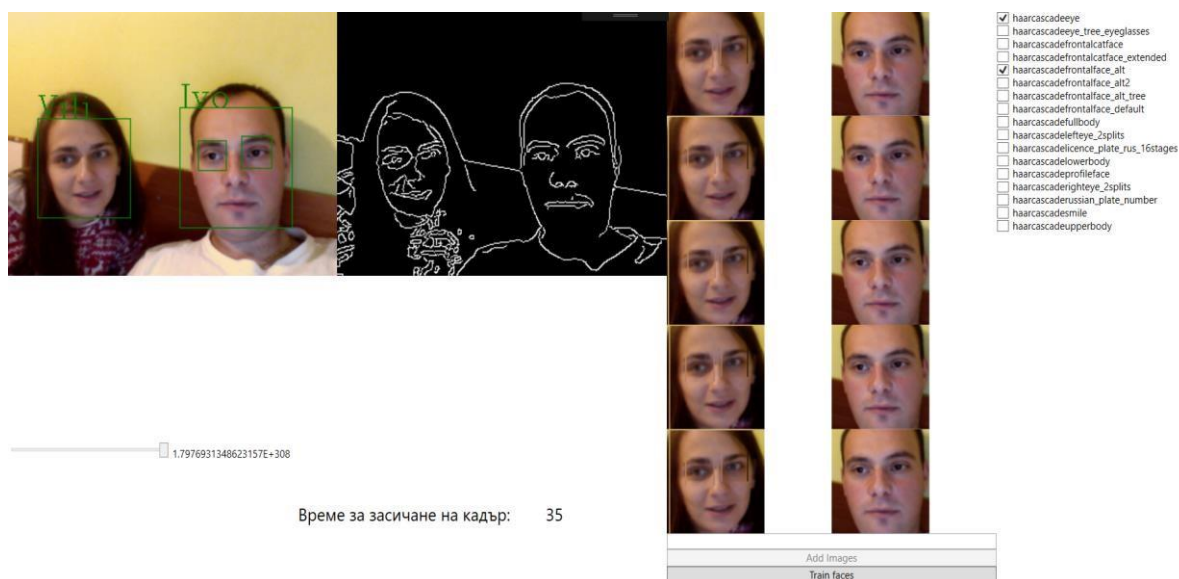
Един от основните тестове, през които трябва да премине подобна система, е следният: колко лица едновременно могат да бъдат засечени и разпознати в реално време. Възможността да бъде засено повече от едно лице е ключова за приложенията, в които може да бъде използвано.

Този тест е проведен чрез увеличаване на броя на хората, които попадат в кадъра на използваната камера, като изводите са следните: приложението успешно засича повече от едно лице едновременно. Същото е в сила и за разпознаването.

На фигура 17 е показана снимка на моментно действие на приложението. В случая са засечени и разпознати и двете лица, налични във видео потока.

В дадения случай са засечени и очите на едното лице (в дясно), докато засичането на очите на лицето вляво се вижда на отделните снети кадри вдясно от видео потока. Очите са разпознати, тъй като освен дескриптор за откриване на лице анфас, е избран и дескриптор за засичане на очи.

Лицата са разпознати правилно на лицата, като имената са изписани над всеки правоъгълник, локализиращ лицето. На главния екран на приложението е изведено поле за *Време за засичане на кадър*, което в конкретния случай е 35 милисекунди за обработка. Засичането на лица се влияе в значително малка от наличната светлина.



Фигура 17. Засичане и разпознаване на две лица

При наличие на повече от две лица резултатите са напълно аналогични.

Функционални възможности за обслужване на множество различни клиенти

Клъстеризацията на системата предоставя възможност за обслужване на повече от един клиент.

При тестовите е симулирана връзка между много машини в клъстерна система. Ако в реални условия клъстерната система е инсталирана на множество машини тя ще работи по-бързо чрез получаване на информация (кадри) от няколко камери, обработването ѝ в реално време и връщане на резултати на всеки съответен свързан към системата клиент.

Тази функция придава изключителни възможности на приложението, а именно едновременно събиране на информация от множество източници и обработването ѝ в реално време. При наличие на повече от един източници, взимането на решения става в реално време и самият коефициентът на разпознаване е по-голям, тъй като наборът от данни, с които се обучава невронната мрежа е по-голям.

Тестове за време при засичане и разпознаване

Тестовите за време се състоят в засичане на времето в милисекунди съответно за засичане на едно лице, за разпознаване на едно лице и за засичане+разпознаване на налично лице.

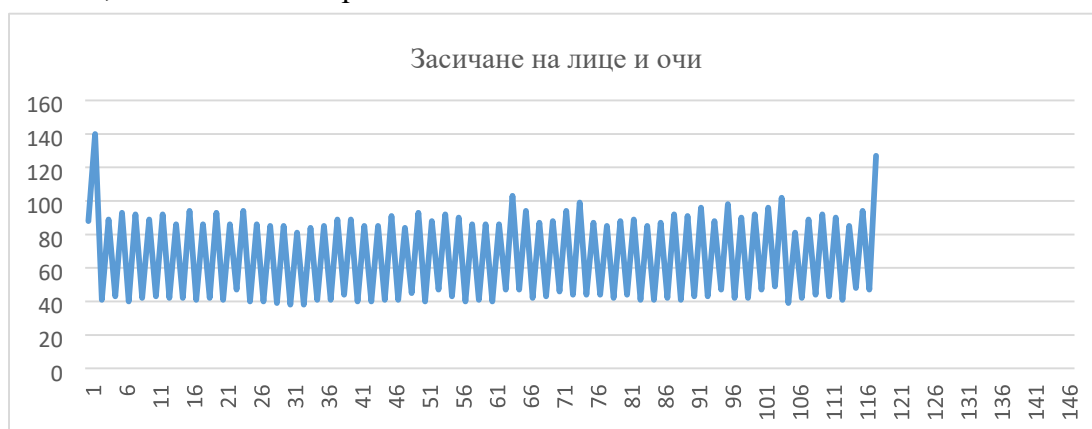
Тестовите са направени с до 145 кадъра, като времето е измерено от системата в реално време.

Резултатите показват, че времето за засичане на лице е в диапазона 40-50 милисекунди (фиг. 18), а за засичане на лице и очи в рамките на лицето - 40-60 милисекунди (фиг. 19).



Фигура 18. Време за засичане на лице, като по абсцисната ос се отчитат брой кадри, а по ординатната – време в милисекунди

Получените резултати са задоволителни при характеристиките на камерата и машината, на която се извършват тестовете.



Фигура 19. Време за засичане на лице и очи, като по абсцисната ос се отчитат брой кадри, а по ординатната – време в милисекунди

Най-важен е фактът, че системата се вписва идеално във времевия диапазон за работа на система в реално време. След като приложението успява да засече лице, а дори и очи в рамките на лицето, за период по-малък от 100 милисекунди, може да се направи изводът, че системата ще засича лице, ако такова е налично, мигновено след стартирането ѝ, като човешкото око няма да отчете забавяне при засичането. При засичане и разпознаване на повече от едно лице системата има нужда няколко десетки милисекунди повече, което показва, че необходимото време не е право пропорционално на броя лица и функционалността за работа в реално време не е нарушена.

Другите два теста за време са за засичане+разпознаване на лице и засичане+разпознаване на лице и очи в рамките на лицето.

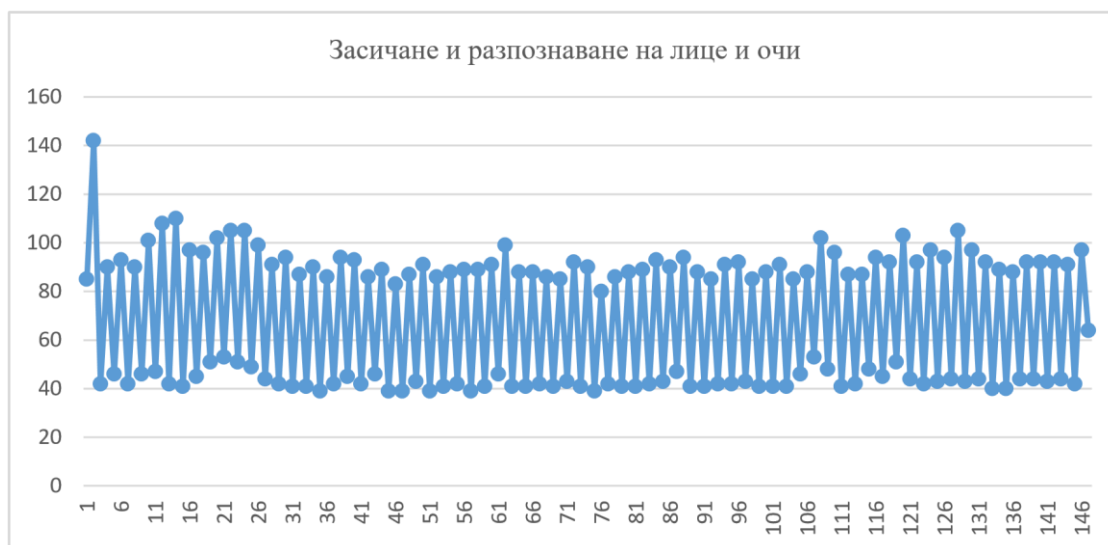
За засичане+разпознаване на лице са използвани 133 кадъра, а при засичането+разпознаване на лице и очи - 147 кадъра.

Процедурата по засичане+ разпознаване на едно лице от системата се изпълнява в рамките средно на 40-50 милисекунди (фиг. 20). Изводът е, че времето за засичане+разпознаване не се различава значително от времето, необходимо само за засичанети на лицето, което е впечатляващ резултат за конкретните тестови условия.



Фигура 20. Време за засичане+разпознаване на лице, като по абсцисната ос се отчитат брой кадри, а по ординатната – време в милисекунди

Необходимото време за засичане+разпознаване на лице и очи заедно е малко поголямо от времето за засичане+разпознаване само на лице и е в диапазона от 40-85 милисекунди, но почти се различава от този при тестовите само за засичане на лице и очи (фиг. 21).



Фигура 21. Време за засичане+разпознаване на лице и очи, като по абсцисната ос се отчитат брой кадри, а по ординатната – време в милисекунди

Резултатите като средно време от тези два теста са изцяло в допустимите рамки за засичане в реално време, като при други условия (пр. повече от едно лица в кадъра) времето ще бъде по-голямо, но със сигурност отново в допустимите граници.

Тестове при различно качество на видео потока

Освен времето за изпълнение на главните задачи за засичане и разпознаване, бързодействието на системата е тествано и при различно качество на получаваните кадри. Това показва устойчивостта на приложението при лошо качество на картината и е показател за минималните изисквания за надеждността му.

Направени са тестове с три различни вида резолюция с различни части от наличното изображение.

При изображение с резолюция 64x48 пиксела се достига време под 1 милисекунда. При следващия тест размерите на кадрите са 160x120 пиксела, като е запазена ¼ от целия кадър. В този случай времето за засичане се увеличава до 5-7 милисекунди и не се наблюдава разлика при разпознаването. При последния направен тест резолюцията е 320x240 пиксела. При този случай се наблюдават промени в бързодействието и на двата алгоритъма, като времето е 20 милисекунди за засичане и отново 20 милисекунди за разпознаване.

Основен извод от тази група тестове е, че не се наблюдава разлика във времето за засичане или разпознаване на лицето въпреки влошеното качество на изображението. Резултатите от проведените тестове са обобщени в Таблица 1. *Таблица 1. Резултати от тестове за време при различно качество на видео потока*

Резолюция	Част от изображението	Време за засичане	Време за разпознаване
64x48 px	1/10	<1ms	<1ms
160x120 px	1/4	5-7ms	5-7ms
320x240 px	1/2	20ms	20ms

Засичане на лице при влошеното качество на видео потока е илюстрирано на фигура 22.

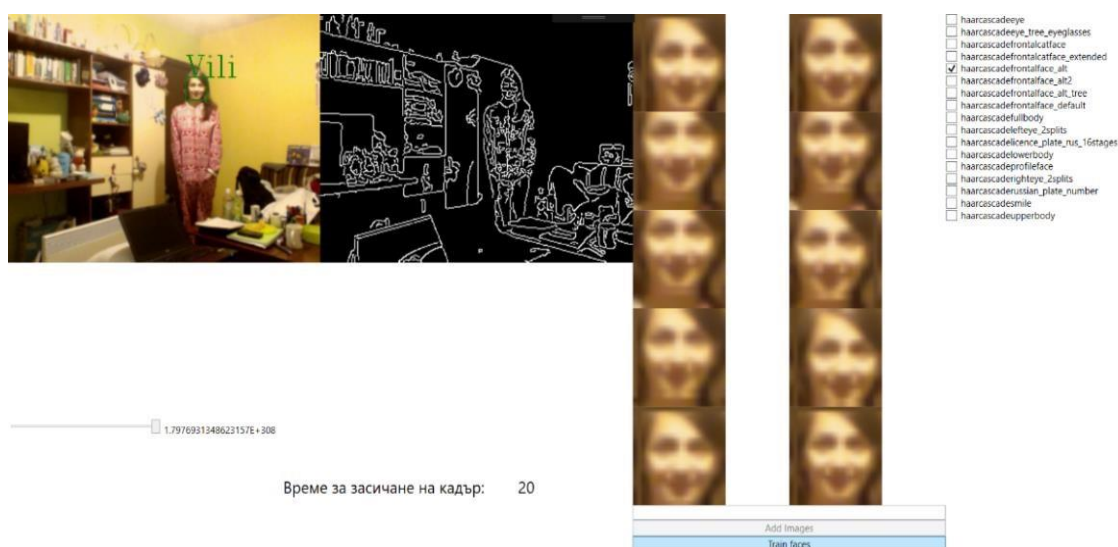


Фигура 22. Засичане на лице при ниско качество на кадрите

Тестове за отдалечено засичане и разпознаване

Системите за засичане и разпознаване на образи се използват най-вече в области, които са свързани с наблюдение на големи групи хора, които не във всички случаи се намират близко до камерата. По тази причина подобна система трябва да бъде устойчива и на разстоянието, на което се засича и разпознава лице.

След направени тестове за засичане и разпознаване при различни разстояния от сканиращата камера, изводите са, че текущата система засича и разпознава до 6 метра при влошено качество на видео потока и до 10 метра при реалната разделителна способност на тестовата камера от 640x480 пиксела. Засичане на отдалечен образ и изображенията на лицето, които са изрязани след процеса на засичане са показани на фигура 23. Качеството е влошено, но времето за засичане и разпознаване не се променя значително.



Фигура 23. Засичане и разпознаване на отдалечен образ

Коефициент на грешно засечени лица и незасечени присъстващи лица

В системите за засичане на образи има два основни критерия: *грешно засечени лица* и *незасечени, но присъстващи в кадъра лица*. При първия критерий алгоритъмът може да засече обект като лице в даден кадър, но това всъщност да не лице. Във втория случай обект, който представлява лице на даден човек, не бива засечено като такова от системата. И двата случая са нежелателни за една система за разпознаване и следователно са предизвикателство за преодоляване.

В текущата разработка са направени няколко оптимизации, в резултат на които се получава по-голям процент на точност при алгоритъма за засичане и оттам и при този за разпознаване на лица. В случая оптимизациите се насочени главно към проблема на грешното засичане, което спомага и за бързината на системата, защото грешно засечените области от даден кадър не се подават на невронната мрежа за разпознаване. Главната оптимизация при търсенето на лице е свързана с намирането на очи в рамките на потенциално открито лице. Построено е следното правило: *наличие на лице* $\leftrightarrow$ *наличие на две несъвпадащи очи*, според което е възможно да има лице само там, където има две правилно позиционирани очи и обратно.

Първо се филтрират намерените правоъгълни области, в които очите не са на правилното разстояние едно от друго или излизат от рамките на дадения правоъгълник с лице в него по следния начин:

1 reference | Ivelin Kirilov Ivanov, 3 hours ago | 1 author, 1 change

```
public static List<Rectangle> FilterFacesRect(this List<Rectangle> faces, List<Rectangle> eyes)
{
    List<Rectangle> rect = new List<Rectangle>();

    foreach (var eye in eyes)
    {
        rect.AddRange(faces.Where(x => x.X <= eye.X && x.Y <= eye.Y).ToList());
    }
    return rect;
}
```

Филтърът, който служи за ограничаване на намерените очи, съдържа ограничителни проверки, свързани със самите разстояния между очите и това дали те се намират на приблизително една и съща координата по ординатната ос:

```

1 reference | Ivelin Kirilov Ivanov, 2 hours ago | 1 author, 1 change
public static List<Rectangle> FilterEyeRect(this List<Rectangle> eyes, List<Rectangle> faces)
{
    List<Rectangle> newEyesList = new List<Rectangle>();

    foreach (var face in faces)
    {
        var temp = eyes.Where(x => x.X > face.X && x.Y > face.Y && x.Right < face.Right && x.Bottom < face.Bottom)
            .OrderBy(x => x.Height)
            .ThenBy(x => x.Width)
            .ThenBy(x => x.Y)
            .ToList();
        if (temp.Count() > 1)
        {
            Rectangle[] leftRightEyes = new Rectangle[2];
            int minYdifference = Int32.MaxValue;
            int maxXdifference = Int32.MinValue;

            for (int i = 0; i < temp.Count; i++)
            {
                for (int j = 0; j < temp.Count; j++)
                {
                    if ((temp[i].Y - temp[j].Y) <= minYdifference
                        && (temp[i].X - temp[j].X) >= maxXdifference
                        && (temp[i].Right < temp[j].X || temp[i].X > temp[j].Right))
                    {
                        leftRightEyes[0] = temp[i];
                        leftRightEyes[1] = temp[j];
                    }
                }
            }
            if (leftRightEyes[0].X != leftRightEyes[1].X)
                return leftRightEyes.ToList();
        }
    }
    return newEyesList;
}

```

Въпреки че тези две филтриращи грешните лица проверки увеличават в известна степен грешките от другия тип (незасечени, но присъстващи лица, при които в дадения кадър присъства лице, което не може да бъде засечено от системата), те се извършват, защото без тях системата би давала по-голям процент грешни резултати. От друга страна грешка от вторият вид може да бъде избегната чрез увеличаване качеството на видео потока, като по този начин ясно ще се отличават детайли като очите и така системата ще работи с висока точност.

V. ЗАКЛЮЧЕНИЕ

В настоящата студия е представена реализирана софтуерна система за засичане и разпознаване на лица в реално време.

Системата обработва изпратеният в реално време от потребителя видео поток, като го разделя на кадри. Тези кадри се трансформират в черно-бели изображения. Следващият етап е подаване на текущите изображения на отделни възли на системата за обработване от невронна мрежа, реализираща алгоритъма ВиолаДжоунс. Накрая потвърдените за реални лица се подават на невронна мрежа, прилагаща принципа на анализ на главните компоненти, за да бъде установена самоличността на лицето. Преди допускане на ново изображение т, невронната мрежа бива обучена с последното състояние на базата данни от изображения за сравнение. След разпознаването, информацията се изпратеща обратно на потребителя. Процедурата по засичане и разпознаване може да бъде изпълнявана за множество различни източници на видео поток едновременно.

Софтуерът, който реализира системата, е написан на програмния език от високо ниво C#, като е използвана библиотеката с отворен код OpenCV.

Възможно бъдещи подобрене на предложената система е възможно с добавяне на хистограма на ориентираните градиенти - алгоритъм, подходящ за локализиране на хора в движение. Друга възможност е прилагане на алгоритъм за прихващане ръбовете на лицето и направата на 3D модел от намерените вектори. Тези подходи са подходящи за подобряване на ефективността на софтуерна система, работеща в реално време, защото предлагат технология за бързо следене на лицето след засичането му и постигането на независимост на засичането и разпознаването от фактори като светлина, ъгъл, разстояние и други.

Библиография и използвани интернет източници

- [1] Orts J., "Face Recognition Techniques," Madison, Wisconsin, 2014.
- [2] West J., "Brief history of face recognition software," Face First, [Online]. Available: www.facefirst.com.
- [3] Goldstein A., L. Harmon, A. Lesk, "Identification of human faces," *Proceedings of the IEEE*, vol. 59, no. 5, pp. 748 - 760, 1971.
- [4] Sirovich L., M. Kirby, "Low-dimensional procedure for the characterization of human faces," *Journal of the Optical Society of America*, vol. 4, no. 3, p. 519–524, 1987.
- [5] Pentland A., M. Turk, "Eigenfaces for recognition," Massachusetts Institute of Technology, Massachusetts, 1991.
- [6] Mayhew S., "History of Biometrics," 2015. [Online]. Available: <http://www.biometricupdate.com/201501/history-of-biometrics>.
- [7] Wechsler H., "Standarts Overview," in *Reliable Face, Recognition Methods, System Design, Implementation and Evaluation*, George Mason, Springer, 2007, pp. 95-98.
- [8] ISO/IEC JTC 1/SC 29, "ISO/IEC 15444-1:2004". USA Patent 800, 1 9 2004.
- [9] Jolliffe T., Principal Component Analysis, second edition, Springer-Verlag, 2002.
- [10] Laplante A., Real-time systems design and analysis, IEEE-Press, 2004.
- [11] Juvva K., "Real-Time Systems," 1998. [Online]. Available: https://users.ece.cmu.edu/~koopman/des_s99/real_time/.
- [12] Milo R., "BioNumbers," Harvard University, 2010. [Online]. Available: <http://www.bionumbers.hms.harvard.edu/>.
- [13] Rauber T., Parallel programming for Multicore and Cluster systems, Springer, 2013.
- [14] Hausdorff F., "Basics of Set Theory, Berlin, 1914.
- [15] Viola P., M. Jones, Robust Real-time Object Detection, Cambridge, 2001.
- [16] Gao W., "On the doubt about margin explanation of boosting," *Science Direct*, vol. 203, no. 2, pp. 1-18, 2013.
- [17] Krishan K., R. Mishra, "Principle Component Analysis," in *Human and face detection and recognition*, Rourkela, 2011, pp. 9-12.
- [18] Jauregui J., Principal component analysis with linear algebra, Philadelphia: Penn Arts & Sciences, 2012.

ОРГАНИЗАЦИЯ УЧЕБНО-ТРЕНИРОВОЧНОГО ПОЛИГОНА ПРОТИВОКОРРОЗИОННОЙ ЗАЩИТЫ СЕЛЬСКОХОЗЯЙСТВЕННОЙ ТЕХНИКИ

*Доцент к.воен.н. Алимов В.В.,
кафедры Материаловедения и технологии машиностроения ФБГОУ ВО
РГАУ-МСХА имени К.А. Тимирязева*

Аннотация: С учетом сезонности работы сельскохозяйственной техники технологии хранения и противокоррозионной защиты становятся актуальными. Формирование профессиональных компетенций у специалистов по применению сельскохозяйственной техники, тракторов и автомобилей включает обучение способам защиты техники от механического и коррозионного воздействия в процессе эксплуатации техники и в нерабочие периоды. Знание обучающимися видов коррозионного воздействия на конструкционные материалы и особенностей коррозии и износа техники в сельском хозяйстве, технологии применения современных средств противокоррозионной защиты позволит значительно сократить потери в сельскохозяйственном производстве. Создание полигона противокоррозионной защиты сельскохозяйственной техники в образовательных организациях высшего и среднего профессионального образования является одним из необходимых условий обучения будущих специалистов инженерных служб, что в конечном итоге позволит снизить затраты на содержание техники на предприятиях АПК.

Ключевые слова: Сельскохозяйственная техника, трактора, автомобили, хранение, эксплуатация, профессиональное обучение, профессиональные компетенции, виды коррозии, противокоррозионная защита, средства защиты, ингибиторы коррозии, полигон, учебные места, учебная группа, подгруппа, технологические операции, консервация, расконсервация, техническое обслуживание, ремонт.

ORGANIZATION OF EDUCATIONAL-TRAINING RANGE OF ANTI-CORROSION PROTECTION AGRICULTURAL MACHINERY

*Alimov V. V. Russian
State Agrarian University - Moscow Timiryazev Agricultural Academy,
Russian Federation, Moscow*

Abstract: Taking into account the seasonality of the work of agricultural machinery, the technologies of storage and anti-corrosion protection become topical. The formation of professional competencies for specialists in the application of agricultural machinery, tractor and automobiles includes training in ways to protect machinery from mechanical and corrosive effects during the operation of machinery and during non-operating periods. Knowledge of the types of corrosive effects on structural materials and the features of corrosion and wear of machinery in agriculture, technology, application of modern anti-corrosion protection means will significantly reduce losses in agricultural production. Establishment of a landfill for the anti-corrosion protection of agricultural machinery in educational institutions of higher and secondary vocational education is one of the necessary conditions for training future specialists of engineering services, which in the end will reduce the cost of maintaining equipment at agricultural enterprises.

Keywords: Agricultural machinery, tractors, automobiles, storage, operation, vocational training, professional competences, types of corrosion, corrosion protection, protective equipment, corrosion inhibitors, training ground, training group, subgroup, technological operations, conservation, re-entry, maintenance, repair

В связи с сезонностью выполняемых разнообразных сельскохозяйственных работ применение сельскохозяйственной техники, тракторов и автомобилей в сельскохозяйственном производстве носит выраженный сезонный характер. Комбайны, навесные и прицепные сельскохозяйственные машины и механизмы работают в году от 150 до 350 часов, остальное время находятся на хранении, причем в большей части на открытых площадках или под навесами.

В то же время стоимость сельскохозяйственных машин и стоимость их эксплуатации достаточно высока.

Так одна из наиболее распространенных в России сельскохозяйственных машин – роторный комбайн Claas Lexion 480 (рис. 1) очень популярна у отечественных сельхозпроизводителей несмотря на то, что машина в Германии снята с производства в 2014 году и предложения по поставкам новых комбайнов отсутствуют. Ценовой диапазон на машины выпуска 2002 – 2004 годов варьируется от 70000 до 80 тысяч евро. В качестве серьезного недостатка отмечается высокая стоимость послегарантийного сервисного обслуживания и отсутствие материалов и информации по самостоятельному ремонту комбайновой техники в полевых условиях. В остальном к данной модели претензий нет, и в хозяйствах она эффективно работает при уборке выращенного урожая. Это один из многих примеров использования бывшей в употреблении качественной импортной сельхозтехники.



Рис. 1. Комбайн Claas Lexion 480 без жатки.

Сельскохозяйственная техника требует обеспечение качественного хранения в межсезонье. В связи с этим одной из основных задач машинно-тракторного парка является создание рациональной организации и технологии хранения машин в нерабочий период. Цель такой технологии – снижение потерь от действия разрушающих факторов:

–климатических – атмосферных осадков, солнечной радиации, воздействия тепла и холода, атмосферного озона, углекислого и сернистого газов и других загрязнений атмосферы;

–обусловленных конструктивным исполнением машин – весовых и монтажных напряжений, вызывающих деформацию деталей и узлов; отсутствие надежной герметизации приводит к проникновению внутрь узлов и агрегатов пыли и атмосферных осадков, вызывающих коррозию и форсированное изнашивание деталей;

–обусловленных спецификой использования машин – воздействием механических напряжений, топлива и масла, пыли и грязи, частиц удобрений и ядохимикатов, продуктов обмолота и др. [1].

Достижение данной цели возможно при высоком уровне профессионального мастерства специалистов по противокоррозионной защите сельскохозяйственной техники, автомобилей и тракторов. Решение данной задачи невозможно без создания без создания учебной материально-технической базы с применением передовых технологий и оборудования, что объясняет необходимость создания в образовательных организациях высшего и среднего профессионального образования учебно-тренировочных полигонов (далее полигон).

В качестве объектов противокоррозионной защиты выбраны сельскохозяйственные машины, автомобили, тракторы, технические средства и оборудование.

Структурная схема полигона противокоррозионной защиты сельскохозяйственной техники (рис. 2).



Рис. 2. Схема полигона противокоррозионной защиты сельскохозяйственной техники.

Полигон предполагает наличие следующих учебно-тренировочных площадок (учебных мест):

- пост очистки и дефектовки машин после окончания выполнения сезонных сельскохозяйственных работ;
- учебное место постановки сельскохозяйственной техники на хранение (консервация);
- пост технического обслуживания сельскохозяйственной техники в период хранения;
- учебная точка – снятие сельскохозяйственной техники с хранения (расконсервация);
- пост ремонта сельскохозяйственной техники;
- учебное место сборки и комплектования сельскохозяйственной техники (новой и после ремонта);
- учебное место разборки списанной сельскохозяйственной техники;
- учебное место обслуживания и противокоррозионной защиты элементов и объектов машинного двора (топливные и масляные резервуары, трубопроводы, элементы складов, погрузочные и разгрузочные механизмы, другая техника и агрегаты, выполняющие работу по обслуживанию сельскохозяйственной техники).

Для качественного проведения занятий группа обучающихся может делиться на подгруппы и через определенные промежутки времени будет производиться перемещение подгрупп на следующее учебное место по схеме.

На учебных местах предусмотрена имитация следующих видов коррозии и неисправностей сельскохозяйственной техники:

- химическая коррозия конструкций сельскохозяйственной техники;
- электрохимическая коррозия;
- газовая коррозия (химическая коррозия в газовой среде при минимальном количестве влаги);
- атмосферная коррозия сельскохозяйственной техники, автомобилей и тракторов;
- подземная коррозия элементов машинного двора (заглубленных резервуаров и трубопроводов);
- биокоррозия (протекающая под влиянием жизнедеятельности микроорганизмов);
- контактная коррозия;
- радиационная коррозия, обусловленная действием радиоактивного излучения;
- коррозия внешним током и коррозия блуждающим током, возникающая под воздействием тока от внешнего источника или под воздействием блуждающего тока;
- коррозия под напряжением, появляющаяся вследствие одновременного воздействия коррозионной среды и механического напряжения;
- коррозионная кавитация, обусловленная одновременным коррозионным воздействием и ударным воздействием внешней среды;
- феттинг-коррозия, появляющаяся в результате вибрации и воздействия коррозионной среды.

Также на учебных местах обучающимся должны быть представлены образцы сельскохозяйственной техники и отдельные конструкции, подвергшиеся коррозионному разрушению различных типов: сплошная коррозия; неравномерная коррозия; избирательная коррозия; местная (локальная) коррозия, проявляющаяся в виде отдельных пятен, язв или точек (питтингов), глубоко проникающих в металл; подповерхностная коррозия; межкристаллическая коррозия; щелевая коррозия [2].

Применяемые на учебных местах консервационные средства, средства противокоррозионной защиты:

- ингибиторы коррозии (маслорастворимые, водорастворимые, летучие ингибиторы и др.);
- консистентные смазки;
- консервационные масла;
- пленкообразующие ингибированные нефтяные составы;
- лакокрасочные покрытия;
- защитные восковые дисперсии;
- бензинобитумные составы;
- другие средства защиты от коррозии.

Также в для защиты от коррозии на учебных местах должны быть представлены пластичные смазки. Пластичные смазки по своим свойствам занимают промежуточное положение между маслами и твердыми смазками, т. е. они сочетают в себе свойства твердого тела и жидкости (при увеличении нагрузки или под действием температуры смазки деформируются и начинают течь, как обычные смазочные масла, но этот процесс является обратимым, т. е. после снятия нагрузки течение смазки прекращается и она вновь приобретает свойства твердого тела) [3].

Смазки предназначены для уменьшения износа, предотвращения задира и заклинивания поверхностей трения, увеличения срока службы деталей механизмов, предотвращения коррозии, герметизации соединений и других функций. С помощью пластических смазок защищаются все металлические изделия и механизмы всех видов, за исключением стальных канатов и случаев, требующих использования консервационных масел или твердых покрытий.

Также для проведения консервационных работ используются отработанные моторные масла.

Обучающиеся в ходе занятий на учебных местах практически выполняют работы по обслуживанию и противокоррозионной защите сельскохозяйственной техники и ее отдельных конструкций с применением различных материалов и технологий.

Выводы по статье

Таким образом, в результате создания в образовательных организациях высшего и среднего профессионального образования учебно-тренировочного полигонов противокоррозионной защиты сельскохозяйственной техники и в ходе поведения практических занятий на учебных местах у обучающимся формируются профессиональные компетенции по выполнению всего комплекса технологических операций в целях обеспечения рациональной организации и технологии хранения сельскохозяйственных машин, автомобилей и тракторов в нерабочие периоды.

Список литературы

- [1.] Пучин Е.А., Гайдар С.М. Хранение и противокоррозионная защита сельскохозяйственной техники. / учебное пособие – М.: ФГНУ «Росинформагротех», 2011, С. 512.
- [2.] Гайдар С.М. Теория и практика создания ингибиторов коррозии для консервации сельскохозяйственной техники. / монография – М.: ФГНУ «Росинформагротех», 2011, 304 с.
- [3.] Сафонов А.С., Ушаков А.И., Гришин В.В. Химмотология горюче-смазочных материалов: Научно-техническое издание / НПИКЦ, 2007. – С. 488