



ТЕХНИЧЕСКИ УНИВЕРСИТЕТ – СОФИЯ  
ФАКУЛТЕТ ПО ТЕЛЕКОМУНИКАЦИИ

---

# ДИПЛОМНА РАБОТА

НА

Александра Иванова Василева, Фак. № 111210082

**Специалност: Телекомуникации**

**Образователно-квалификационна степен:** бакалавър

**Тема:** Разпознаване на движения и жестове от изображения и видео последователности

**Научен ръководител:**

гл. ас. д-р Агата Манолова ТУ-София

**Консултант:**

гл ас. д-р инж. Иво Драганов ТУ-София

София 2014

В настоящата дипломна работа ще се разгледа разпознаването на движения от изображения и видео последователности, като се използват данните на Weizmann.

Разпознаването на човешката активност е представено в три нива на фиг.1 от първа глава. Също така са описани подробно някои методи за извличане и представяне на признаци за разпознаване на човешки движения. Проследяването на състоянието на човешкото тяло, неговото движение в пространството е приложимо в сферите на наблюдение, забавление и здравеопазване.

# I. Глава

## 1. Въведение.

През последните години, автоматично разпознаване на човешката дейност привлича вниманието в научната сфера за видео анализ, породено от увеличаващото се търсене на множество приложения за наблюдение, забавление и здравеопазване.

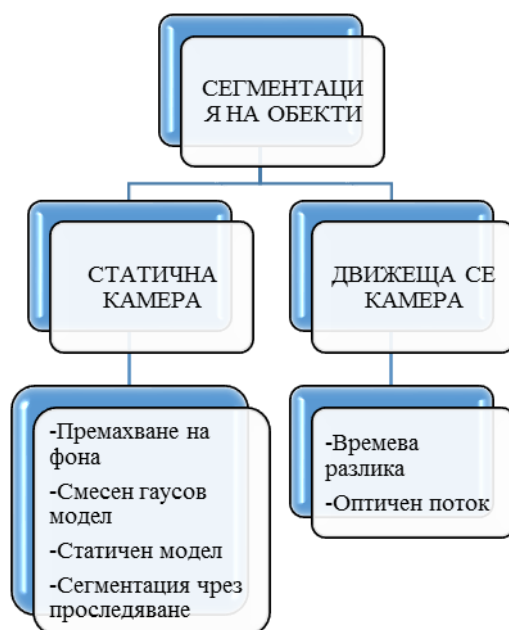
Разпознаването на човешката активност може да бъде представено в три нива: ниско ниво технология на ядрото, средно ниво системи за разпознаване на човешката активност и високо ниво приложения както е показано на фиг.1.

Фиг.1 Обща система за разпознаване на човешка активност



В първия етап на технологията на ядрото, сегментацията на обект се извършва във всеки фрейм от видео последователността, за да се извече желанния обект. В зависимост от мобилността на камерата, сегментацията на обекта може да бъде категоризирана като два типа сегментация: сегментация от статична камера и сегментация от движеща се камера, показани на фиг.2.

Фиг. 2 Категории за сегментация на обекти



При сегментацията от статична камера, камерата е фиксирана на специална позиция с фиксиран ъгъл. Следователно изгледът към обекта и фона не се променя. Най-използвания метод за сегментация от статична камера и премахването на фона [1-3], защото е много лесен и ефективен.

От друга страна, при сегментация от движеща се камера, камерата се движи с фотограф или робот. Този вид сегментация е много по-голямо предизвикателство от сегментацията от статична камера, защото в допълнение към движението на обекта, трябва да се вземе предвид движението на камерата и промяната на фона.

Във вторият етап на технологията на ядрото, характеристиките на сегментирания обект като форма, силует, цвят и движения са извлечени и представени от признаци. Признаците могат да бъдат категоризирани в четири групи: пространствено-времеви обем, честота, локални дескриптори и моделиране на тялото, както е показано на фиг.3.

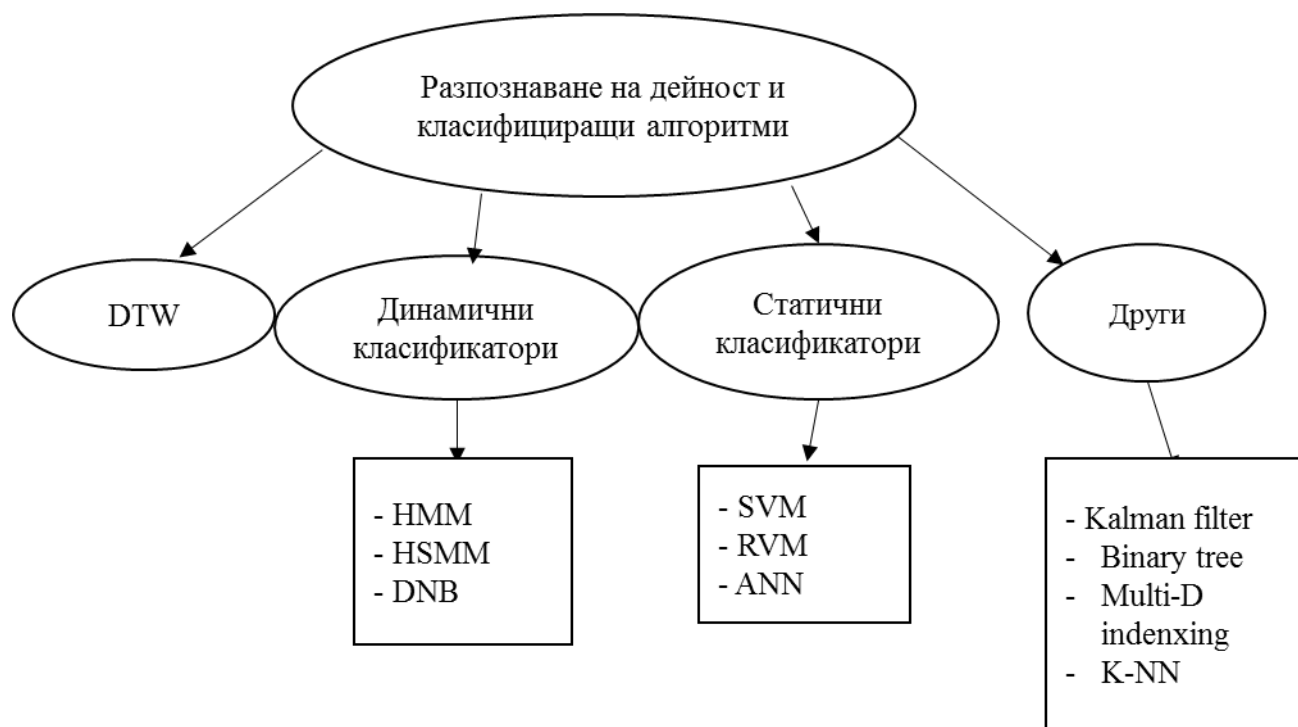
Фиг.3 Категории за извличане и представяне на признаци



Пространствено-времевият обем (STV) [4] се построява като признаково изображение чрез конкатенация на поредни силуети на обекта по времевата ос. Информацията от честотния обхват също може да се използва, по конкретно дискретната Фурие трансформация (DFT)[5], която е широко използвана за представяне на геометрична информация за структурата на обекта. Локалните дескриптори STIF [6,7] и HOG [8] ще бъдат разгледани малко по подробно в точка 3 от тази глава. Въпреки всичко горе споменатите методи не могат да покрият напълно движението на цялото тяло. Поради тази причина са разработени някои методи за моделиране на човешкото тяло [1, 9]: прости петна (simple blobs), 2D и 3D модели.

В третият етап на технологията на ядрото, откриването на дейност и класифициращите алгоритми се използват за разпознаване на различните човешки дейности, базирани на съответните признаци. Те могат да бъдат категоризирани като dynamic time warping (DTW), динамични класификатори, статични класификатори и други, както е показано на фиг. 4.

Фиг. 4 Категории за разпознаване на дейност и класифициращи алгоритми



The dynamic time warping (DTW) [10, 11], метод за измерване на подобие между две времеви последователности, които могат да варират по време или скорост, е един от най-често срещаните времеви класифициращи алгоритми поради своето лесно изчисление. Обаче, DTW не е подходящ за голям брой класове с много вариации. От динамични класификатори най-известни са скритите Маркови модели (HMM) [1] и Динамични Бейсови мрежи (DBN) [13]. Статичните класификатори като Машини за поддържащи вектори (SVM) [14], Relevant Vector Machine (RVM) [15] и Изкуствена невронна мрежа (ANN) [16], също могат да се използват в този етап. Други класифициращи алгоритми са филтър на Калман [17], binary tree [18], multidimensional indexing [19], и К-близкия съсед (K-NN) [5].

## 2. Сегментация на обекти

За разпознаване на човешка дейност, първият етап е да се направи сегментация на обекта, т.е. човешките обекти са сегментирани от фоновото изображение. Въз основа на подвижността на камерата, сегментацията на обекти може да бъде разделена в две категории, сегментация от статична камера и сегментация от движеща се камера.

### 2.1. Статична камера

При сегментация със статична камера, камерата е фиксирана на специална позиция и ъгъл. Тъй като фона никога не се движи, може да се изгради модел, чрез който обектите на преден план могат да бъдат лесно сегментирани от изображението.

### 2.1.1. Премахване на фона

Най-често срещания метод при сегментация със статична камера е премахването на фона, понеже е лесен и ефективен. Моделът на фона съдържа само стационарна сцена без никакви обекти на преден план и всяка промяна в изображението се счита, че се дължи на движещ обект. Следователно обекта от преден план може да бъде получен чрез изваждане на текущото изображение от фоновото, последвано от праг по амплитуда, за да се получи маска за сегментация. Най-общо казано, премахването на фона е лесно и ефективно, но това понякога води до неточна класификация на пикселите. Непрекъснатото и ефективно осъвременяване на фона като отговор на постепенната промяна в него също така поставя някои предизвикателства.

### 2.1.2. Смесен Гаусов модел (GMM)

Вместо простия модел една Гаусова функция за пиксел, стойностите на пиксела в позиция  $(x, y)$  може да бъде моделиран с по сложна форма, като смесица от Гаусови функции, която се пригажда към различни сценарии на фона. Смесеният Гаусов модел (GMM) се прилага много широко в различните сфери, за да позволи приспособяване към мулти-модални среди. Най-общо, GMM се обучава чрез алгоритъм за максимизиране на математическото очакване. Колкото е по голяма вероятността за дадена стойност на пиксела при GMM, толкова е по вероятно дадения пиксел да принадлежи на фона, т.е. за една последователност от изображения един пиксел от изображение се класифицира като принадлежащ на обекта от преден план, ако вероятността на неговата стойност е по-малка от предварително определен праг.

## 2.2. Движеща се камера

За разлика от статичната камера, движещата се е с променящо се местоположение и ъгъл. Сегментацията от движеща се камера е по-голямо предизвикателство, защото два въпроса трябва да бъдат разгледани едновременно: движението на фона и движението на всеки обект от преден план. Обикновено необходимо е разделяне на движението на камерата от това на обектите.

### 2.2.1. Времева разлика

Най-използвания метод при сегментация от движеща се камера е времевата разлика между последователни кадри. За разлика от сегментацията от статична камера, където фона е стабилен, фонът при движещата се камера се променя постоянно, за това не е удачно да се изгражда модел за фона. Вместо това движещият се обект се открива, като се вземе разликата на последователните кадри  $t-1$  и  $t$ . Въпреки това движенията на камерата и движенията на обекта се смесват в движещата се камера. Следователно движението на камерата се оценява на първо място.

### 2.2.2. Оптичен поток

Друг метод за сегментация от движеща се камера е оптичният поток, който обозначава преместването на една и съща сцена в последователност от изображения в

различни времеви моменти. Оптичният поток за дадена последователност от изображения може да бъде грубо оценен чрез алгоритъма на Lucas-Kanade-Tomasi (LKT) [20], който използва информацията за пространствен интензитет, за да насочи търсенето към позиция, която дава най-доброто съвпадение.

### 3. Извличане и представяне на признаци

Вторият етап за разпознаване на човешка дейност е извличане и представяне на признаци, където важните характеристики от графичните кадри са извлечени и представени по систематичен начин, като признаци. Извличането и представянето на признаци има решаващо влияние върху степента на разпознаване, т.е. е особено важно да се изберът признаци на кадрите по подходящ начин. Във видео последователност признаците, които улавят пространствената и времевата връзка са изнесени като пространствено времеви обеми (STV)[4]. Глобалните признаци се определят от цялото изображение, но те са чувствителни на шум, оклузии и вариации на гледната точка. За това вместо да се използват глобални признаци, някои методи предлагат използването на локално извлечени признаци, които са по-малко чувствителни на шум, оклузии и в някои случаи на ротация и промяна в мащаба. Освен глобални и локални признаци някои методи предлагат използване на моделиране на човешкото тяло, които се използват да се оцени позата на тялото и да се проследят отделните негови елементи.

#### Локални дескриптори

Локалните дескриптори са по-малко чувствителни на шум, оклузии и в някои случаи на ротация и промяна в мащаба, като scale-invariant feature transform (SIFT)[6], хистграма на ориентираните градиенти (HOG) [8], nonparametric weighted feature extraction (NWFE) [21], Lucas-Kanade-Tomasi (LKT) [20], основани на формата [15] и основани на външния вид [22].

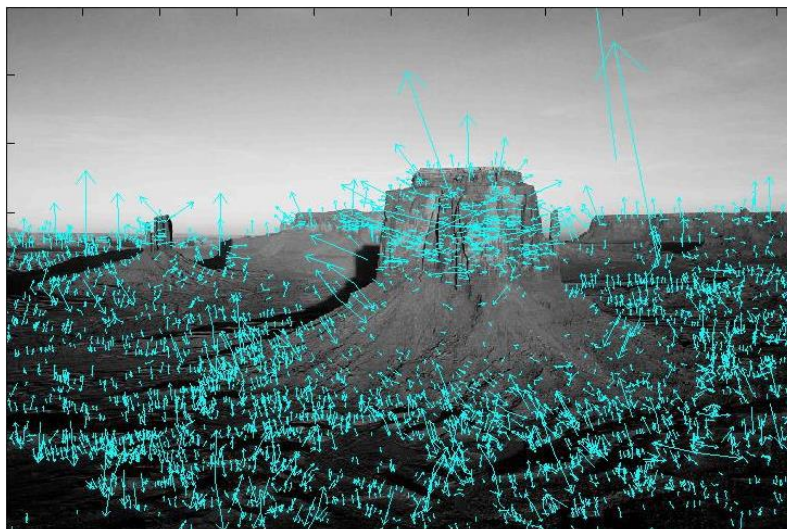
#### -SIFT

Lowe [7] първи предлага клас от локални признаци на изображение, които са инвариантни на мащабиране, трансляция и ротация и частично инвариантни на промени в осветеността или афини проекции. Тези признаци са познати под името scale invariant feature transform (SIFT). Както е предложено в [7], четирите основни стъпки приложени върху едно изображение, за да се генерират SIFT признаци са следните. Първа стъпка чрез разлика на Гаусови функции (difference-of-Gaussian(DoG) functions) се търси спрямо всички мащаби и позиции на изображението, за да се определят точки на интерес, които са инвариантни на мащаб и ориентация. Втора стъпка е локализация на тези ключови точки, основавайки се резултатите от първа стъпка се използва детайлен модел, за да се определи мащаба и положението и след това по-стабилните точки се избират за ключови. Третата стъпка се основава на посоки на градиента в изображението. Последната стъпка е дескриптор на тези ключови точки. За всяка ключова точка се изчисляват градиентите на изображението в даден мащаб и след това се трансформира в представяне, използващо локална промяна във формата и осветеността. Един пример за избиране на ключови точки за SIFT признаци е показан на фиг.5 [7]. Обаче SIFT



използват само полутоновата информация, т.е. цвета не играе роля при тяхното изчисляване

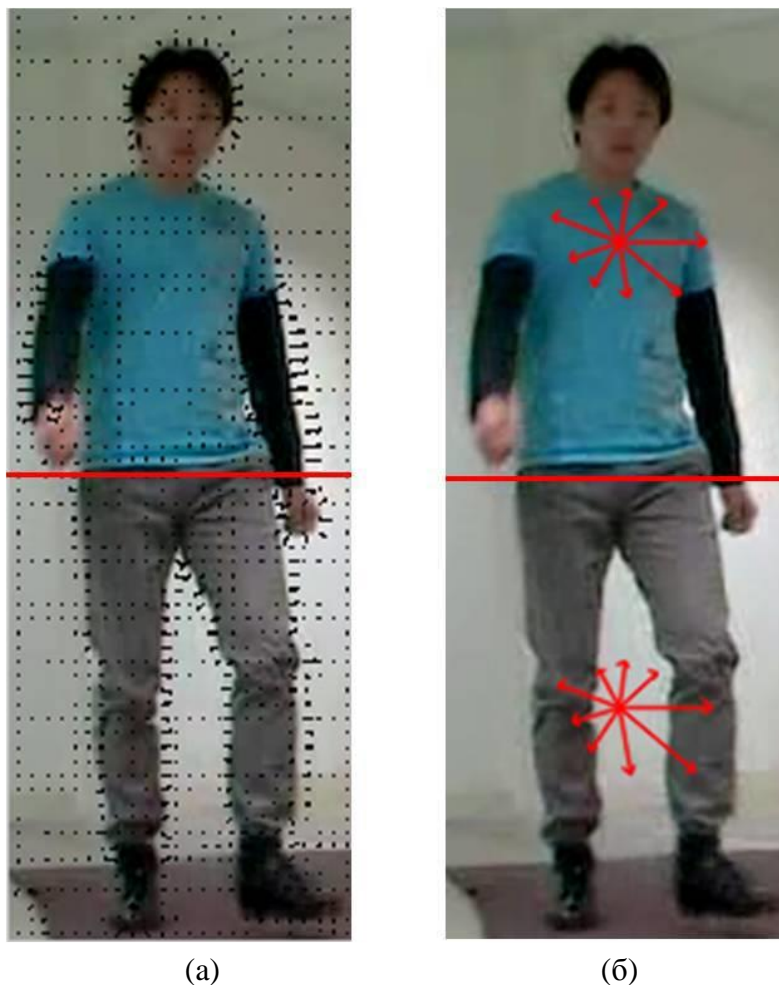
Фиг. 5. 2D SIFT дескриптор, където стрелките показват посоките на локалния градиент на изображението, и информация за мащаб, ориентация и позиция [7].



#### -HOG

Dalal and Triggs [8] са първите, които предлагат дескриптор чрез хистограма на ориентиран градиент (histogram of oriented gradient (HOG)) за намиране на хора. HOG се базира на нормализирани локални хистограми на ориентацията на градиента в изображението. Пример за HOG дескриптор е показана фиг. 6. Една от негативните страни на признаците HOG е, че те са локални признаци, изчислени при фиксиран мащаб, т.е. размерът на човека в изображението може да има голямо влияние върху резултата от разпознаването.

Фиг.6. Пример за HOG дескриптор: (а) изображение на градиентите в локална решетка; (б) HOG дескриптор с девет бина за горната и долната част на тялото [8]



-Признаци основани на външния вид

За сравнение с признаци базирани на формата, признаците базирани на външността дават повече отличителна информация като цвят и са по-стабилни при оклузии. Няколко различни вида дескриптори базирани на външността са предложени в литературата [22] като histogram of local appearance context (HLAC) за определяне на 3D позата и дескриптори за local appearance context (LAC).

#### 4. Откриване на дейност и класифициращи алгоритми

След избирането на подходящи признаци от изображение или видео, откриването на дейност и класифициращи алгоритми е следващият етап, който се взема под внимание при разпознаването на човешка активност. За да се постигне по-добра степен на разпознаване, от съществено значение е да се избере подходящ класифициращ алгоритъм, използвайки избраното признаково представяне. Един от най-известните класифициращи алгоритми е dynamic time warping (DTW) [23], който е мярка за подобие на две последователности, който е възможно да са с различни дължини и с честота. По принцип алгоритмите могат да бъдат разделени на динамични и статични класификатори.

#### 4.1. Dynamic time warping (DTW)

DTW [23] е един от динамичните алгоритми, който измерва подобие (разстояние) между две последователности, т.е. един вид шаблон за подобие между последователности. Предимството на DTW е, че бърз и лесен, но може да изисква допълнителни шаблони за различните ситуации, т.е. изчислителната цена съвпада с този голям брой шаблони.

#### 4.2. Динамични класификатори

Един от най-известните динамични класификатори е скрития Марков модел (hidden Markov model (HMM)) [24].

Скритият Марков модел (HMM) е статистически модел, при който системата, която е била моделирана, се приема да бъде Марков процес с неизвестни параметри и целта е да се определи скритият параметър от изследваните параметри. Извлечените параметри по модела могат после да бъдат използвани за по-нататъшни анализи. Всеки определен (наблюдаван) 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. HMM може да бъде разглеждан като най-обикновена динамична Байесова мрежа.

Скритите Маркови модели са особено известни с приложенията си в разпознаването на глас, ръкопис, жестове, част от реч, музикална тема и др

#### 4.3. Статични класификатори

В сравнение с динамичните класификатори, статичните не могат да бъдат използвани за генериране на подобие с условна вероятност  $P(X,Y)$ , тъй като дължината на оригиналната последователност варира и трябва да бъде трансформирана във фиксирана дължина. Примери за статични класификатори са машини за поддържащи вектори (SVMs) [14], relevance vector machines (RVMs) [25] и изкуствени невронни мрежи (ANNs) [26].

##### - Машина за поддържащи вектори (SVM)

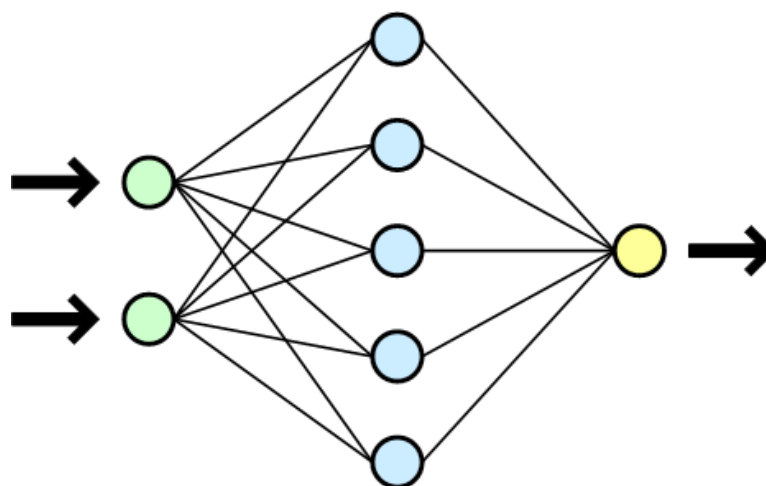
Машина за поддържащи вектори (SVM) [14] е един от най-популярните, основани на разделяща равнина, класификатори с обучител. Използва се да раздели данните в два класа. Целта на SVM е да намери оптималната равнина, като максимизира полето между двата класа. Точките на границата на разделителната равнина се наричат поддържащи вектори. Основната негативна страна на SVM е по-високата изчислителна тежест, която е необходима при обучаващата фаза.

##### - Изкуствена невронна мрежа (ANN)

Математическият аналог на биологичната невронна мрежа представлява множество от взаимно свързани прости изчислителни елементи (неврони). Всеки неврон приема сигнали от другите (под формата на числа), сумира ги, като сумата минава през

активационна функция, и така определя своята активация, която се предава по изходящите връзки към другите неврони. Всяка връзка има тегло, което умножавайки се със сигнала, определя неговата значимост (сила). Теглата на връзките са аналогични на силата на синаптичните импулси, предавани между биологичните неврони. Отрицателна стойност на теглото съответства на подтикващ импулс, а положителна - на възбуждащ. В невронната мрежа обикновено винаги съществуват входен и изходен слой от неврони, във входния се въвежда информацията към мрежата, след това сигналите от входните неврони преминават през един или няколко слоя от междинни (скрити) неврони, според топологията на невронната мрежа, като сигналите накрая стигат до изходния слой, откъдето се чете получената информация.

Фиг.7 Изкуствена невронна мрежа



#### 4.4. Други

Сред класифициращите алгоритми за човешка активност: DTW, динамичните класификатори, като HMMs, и статичните класификатори, като SVMs, RVMs и ANNs, има също така много други техники, които могат да бъдат използвани, като филтър на Калман [17], binary tree [18], multidimensional indexing [19], и К-близкия съсед (K-NN) [5]

##### - К-близкия съсед (K-NN)

Класифициращ алгоритъм основан на принадлежност към клас, към който принадлежат К най-близки съсед. Точката/векторът е класифициран към класа, който е най често срещан измежду К най-близките съсед. К-NN е прост метод с малко параметри. Изчислителното време за тестовата фаза е независимо от броя на класовете. Основната негативна страна на К-NN е, че класификацията е чувствителна към избора на К съседите.

## II. Глава

В тази глава ще се представи теоритично решение на настоящата инженерна задача. Избраният подход за разпознаване на движение от видео последователности се базира на направения обстоен литературен анализ в първа глава. Той се състои от две фази: първа обучаваща фаза и втора тестова фаза.

В обучаващата фаза ще бъдат конструирани Motion history image (MHI) на видео последователност и ще се изчисли Structure similarity index measurement (SSIM), като направи предварителна обработка на видео последователността.

Отделните етапи на всеки един алгоритъм ще бъдат описани детайлно в настоящата глава. По време на тестовата фаза за непозната видео последователност ще бъде класифицирано движението на движещия се обект, като класовете се делят на: навеждане, скачане с махане на ръце, скачане, скачане на място, тичане, скачане на страни, скачане на един крак, ходене, махане с една ръка и махане с две ръце.

### 1. Motion history image (MHI)

#### 1.1. Някои важни характеристики на MHI:

- MHI може да представи последователност от движение по компактен начин. В този случай, последователността от силуети се преобразува в полу-тоново изображение, където се съхранява информацията за главното движението.

- Поради същността на алгоритъмът, MHI може да се създаде и реализира в условия на ниска осветеност, където структурата не може лесно да бъде открита по друг начин. Например, в приложения за наблюдение е много ефективен.

- Представянето на MHI не е толкова чувствително към шумове около силуета, дупки, сенки и липсващите части [27]. Въпреки това, ако има значително присъствие на сянка в движещата се сцена, MHI също включва важната информация за сянката.

- Полутоновото MHI е чувствително към посоката на движение, защото демонстрира посоката на потока на движението. Ето защо е по-подходящо за разпознаване на действия, извършени в противоположни посоки. Например, действието в лявата посока е по-различно от същото в дясната посока [28].

- Пази история на времевите промени на мястото на всеки пиксел, който след това се разпада с течение на времето.

- MHI изразява поток на движение или последователност, с помощта на промяната интензитета на всеки пиксел във времето. MHI представя в един кадър (изображение) цялата времева информация.

#### 1.2.Конструкция на Motion History Image

Интензитет на всеки пиксел в MHI е функция на плътността на движение в това място. От изображението на MHI, можем да забележиме, че по-скорошното движение е по-ярко. Приемаме, че две съседни или последователни изображения могат да бъдат описани от:

$$\begin{aligned} I(x, y, t) &= b_t(x, y) + m_t(x, y) + n_t(x, y) \\ I(x, y, t + 1) &= b_{t+1}(x, y) + m_{t+1}(x, y) + n_{t+1}(x, y) \end{aligned} \quad (1)$$

където  $b_t(x, y)$ : Статичен фон за  $t_{th}$  кадър.  $m_t(x, y)$ : Движещ се обект за  $t_{th}$  кадър.  $n_t(x, y)$ : Фонов шум за  $t_{th}$  кадър.

Сега, ако вземем предвид подхода за разлика между последователни кадри за извличане на движещи се обекти, можем да получим,

$$\begin{aligned} diff(x, y, t) &= I(x, y, t + 1) - I(x, y, t) \\ diff(x, y, t) &= b(x, y) + md(x, y, t) + nd(x, y, t) \end{aligned} \quad (2)$$

където  $b(x, y, t)$ : Препокрита област в последователни кадри.  $md(x, y, t)$ : Област на движение.  $nd(x, y, t)$ : Шум (който може да бъде игнориран в зависимост от ситуацията).

Имайки предвид, че  $diff(x, y, t)$ , съдържа част от движещия се обект, аберацията в фона, която се дължи на движението и шума, което довежда до неправилни резултати и движение с ниска скорост, не може да бъде лесно се открито. За да се намалят тези пречки, може да се конвертира това полу-тоново изображение  $diff(x, y, t)$  в двоично изображение,  $diff^c(x, y, t)$ . Затова, се дефинира  $D(x, y, t)$ , като,

$$D(x, y, t) = [diff(x, y, t) * \tau] / 255 \quad (3)$$

След това, чрез наслагване на последователностите  $D(x, y, t)$ , МНІ може да бъде възпроизведено. МНІ може да бъде генерирано с помощта на разлика от кадри (difference of frames (DOF),

$$\Psi(x, y, t) = \begin{cases} 1 & \text{if } D(x, y, t) > \xi \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

където  $\Psi(x, y, t)$  е бинаризацията на разликата от кадри чрез определяне на праг  $\xi$ .

Параметърът  $\xi$  е минимална разликата в интензитета между две изображения за откриване на промяна. DOF на  $t_{th}$  кадър с разлика в разстоянието  $\Delta$  е

$$D(x, y, t) = |I(x, y, t) - I(x, y, t \pm \Delta)| \quad (5)$$

Сега, можем да изчислим МНІ ( $\Psi(x, y, t)$ ) от горната осъвременена функция ( $H_\tau(x, y, t)$ ) по рекурсивен начин,

$$H_\tau(x, y, t) = \begin{cases} \tau & \text{ако } \Psi=1 \\ \max(0, H_\tau(x, y, t - 1) - 1) & \text{в противен случай} \end{cases} \quad (6)$$

където  $(x, y)$ : Позицията на пиксела.  $\tau$ : Време, което определя последователните кадри.  $\Psi(x, y, t)$ : Осъвременена функция, която определя наличие на движение.  $\tau$ : Решава времевата продължителност на МНІ.

По този начин, можем да имаме изображение със скаларни стойности (т.е., МНІ изображение), където пикселите, принадлежащи на движещ се обект са по-ярки. Горното уравнение се основава на [29], където стойностите на пикселите са намалени с единица (т.е., decay параметър е 1) ако тези пиксели не принадлежат на движещият се обект. Например, ако  $I(x, y, 2)$  - кадър 2 няма движещ се обект и  $I(x, y, 1)$  е равно на  $I(x, y, 2)$ . Тогава резултатното МНІ ще бъде 100% черно или всички пиксели ще бъдат нула. Ако  $I(x, y, 3)$  има движещ се обект в позиция на пиксели  $(2,3)$ ,  $(3,3)$  и  $(3,4)$ , тогава тези пикселни стойности (които са били нула до сега) ще бъде заменени със стойността на номера на видео кадъра от тази сцена или 255, която е максималната стойност за полутоново изображение, а останалата част от пикселите, който не принадлежат на движещият се обект ще останат нула.

Съгласно горното уравнение, ако няма движеща се сцена в следващия кадър  $I(x, y, 4)$  стойностите на пикселите  $(2,3)$  и  $(3,4)$  -новите стойностите на пикселите за  $(2, 3)$  и  $(3,4)$  ще бъде  $H_t(x, y, t-1)-1$ , т.е. предната стойност ще бъде намалена с "1". Така че, ако вземем максималната стойност 255, в този случай, новата стойност ще бъде  $(255-1) = 254$ . Въпреки това, новата стойност за пиксел  $(3,3)$  ще остане същата. По този начин можем да намерим изображение, където минималната стойност на пиксела ще бъде "0", което означава, че тези пикселни стойности с "0" нямат движеща информация или тя е намалена на "0", тъй като дълго време не е имало движение.

Вместо намаляване на стойността с "1", когато няма движеща се информация в този конкретен пиксел, след като е имало информация за движението преди, може да се намери ситуации, при които може да бъде необходимо да се намалят повече стойности на пиксели всеки път когато няма движеща се сцена. Ако се определи параметъра разпад на МНІ, може да се запише:

$$H_{\tau}(x, y, t) = \begin{cases} \tau & \text{ако } \Psi=1 \\ \max(0, H_{\tau}(x, y, t-1) - \delta) & \text{в противен случай} \end{cases} \quad (7)$$

Стойността на параметъра на разпад ( $\delta$ ) може да бъде 1 или повече.

Актуализираната функция ( $\Psi(x, y, t)$ ) е рекурсивно изжикана за всеки кадър или поредица от последователни кадри, които са анализирани в това видео. Окончателното МНІ изображение записва с течение на времето историята на движение в него. Можем да вземем окончателния МНІ шаблон като  $H_{\tau}(x, y, t)$ . Може да се наложи прилагането на средно филтриране, за да се изглади МНІs и премахване на „salt-pepper“ шум или прилагане на Гаусов филтър.

Фиг. 8. Летяща в небето птица: съответното МНІ (средното изображение) и модифицирано МНІ (дясно изображение). Очевидно е, че и двете представяния могат да съхраняват информацията на потока на движение или история. С любезното разрешение от Springer Science+Business Media B.V.—from Meng et al. [30]



## 2. Motion Energy Image (MEI)

### 2.1. Някои важни характеристики на MEI

-MEI осигурява само областите, където присъстват движение или движещи се сцени. С МНІ, от друга страна, показва, къде е настъпило движението или в която посока и интензивност.

-От MEI, можем лесно да се визуализираме видео последователност, където е „пометено“ от определена област на изображението.

- Може да се определи умело формата на движещите се области във видео. Въпреки това, понякога поради наличие на шум, изображението MEI може да покаже някои части, които не трябва да се разглеждат като движещи се области.

- MEI е изображение, което показва, областта на настъпване на движение.

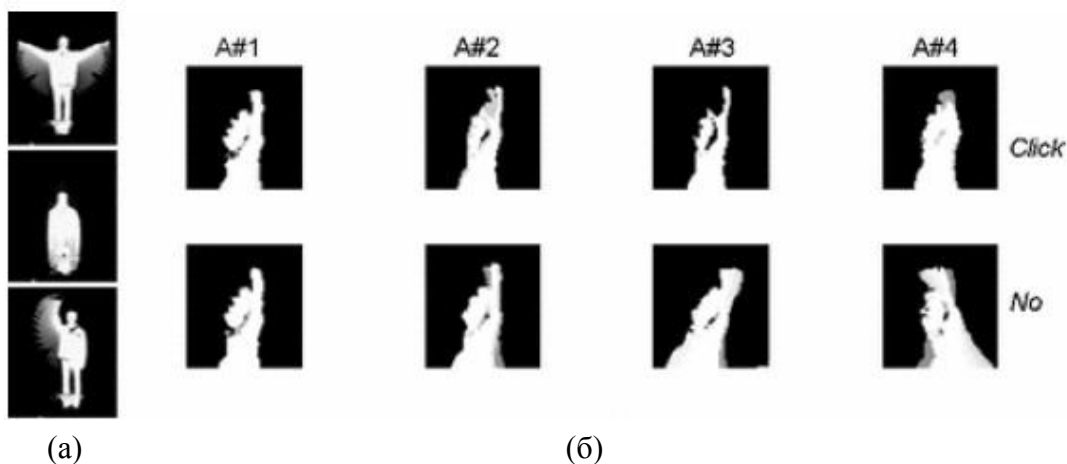
- Двоичното MEI първоначално се изчислява като индекс в библиотеката „Действие“. Това грубо описва пространственото разпределение на енергията на движението за дадена цел на дадено действие.

### 2.2. Конструкция на Motion Energy Image (MEI)

Motion Energy Image (MEI), което обхваща областите на движение на видео сцена. MEI е бинарно изображение, в което бялата област представлява срещащите се движения, а черната област е областта, където няма движения. Първоначално MEI се споменава като Binary Motion Region (BMR) от [31]. Motion Energy Image (MEI) се определя въз основа на пространствен модел на движение на една сцена като бинарно изображение. MEI представлява изображение, където е настъпило движение във видео сцена.



Фиг. 9. (а) МНІ изображение за три събития. (б) МНІ признаци, изчислени с помощта на четири действия на събитията Click and No. С любезното разрешение от Springer Science+Business Media B.V.—from Alahari and Jawahar [32], Fig. 3, Springer, 2006



Фиг. 10. Примери за МНІ и модифицирано МНІ за шест действия от КТН набор от данни: Най-горен ред за кадри за действие; среден ред изобразява съответните МНІ изображения; и долния ред показва модифицирано МНІ. С разрешение от Springer Science Business Media BV-от Meng et Al. [33]



Ако две последователни изображения имат някои промени, т.е. имат някакво наличие на движение, тогава конструираното МНІ изображението ще има пикселни стойности "1" за тези пиксели, където налице са промените. По същия начин, ако следващото изображение има някакви промени пиксел или движеща се част, след това допълнителните пиксели ще се променят от "0" на "1". Чрез този начин, кумулативното бинарно изображение представя историята на присъствието на движение като бинарно изображение. Белите пиксели имат пикселна стойност "1" и могат да се отбележат като енергията на това изображение.

Според [34], нека да приемем  $I(x, y, t)$  да бъде последователност от изображения и нека  $D(x, y, t)$  е последователност от бинарни изображения, които показват областите на

движещите се сцени във видео последователността. Това може да бъде направено чрез използване на метода изваждане на фона, ако фона е наличен или ако фоновия модел може да се развива в движение. Други известни опции са кадрово изваждане или оптичен поток, и т.н.  $D(x, y, t)$  се генерира от един от тези подходи. След това бинарното MEI може да бъде определено от,

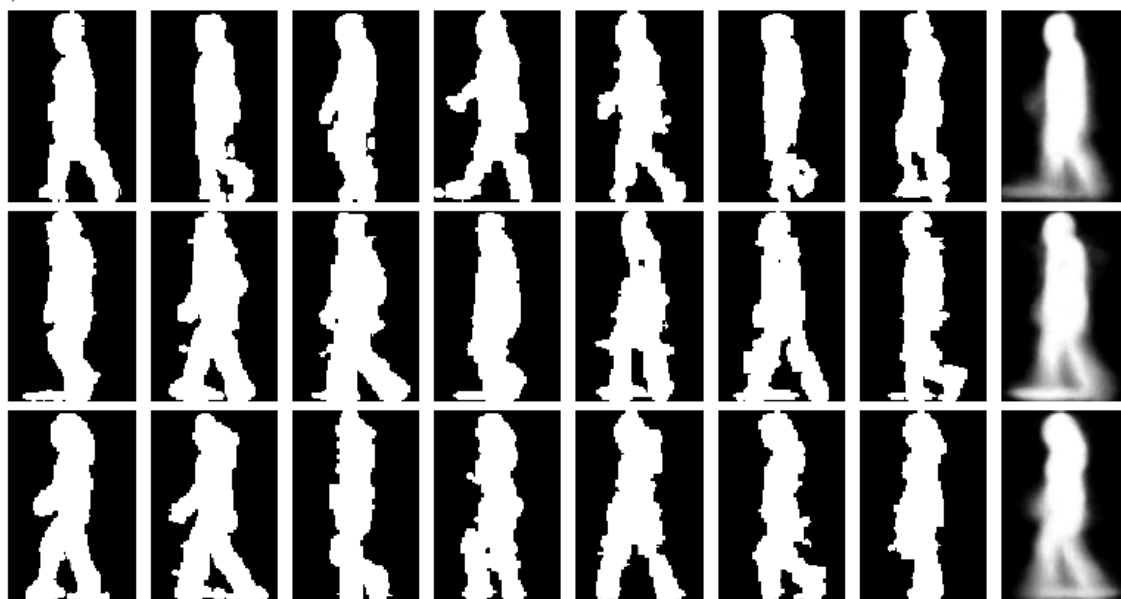
$$E_{\tau}(x, y, t) = \bigcup_{i=0}^{\tau-1} D(x, y, t - i) \quad (8)$$

В горното уравнение, критичен параметър е изборът на  $\tau$ . Това е от решаващо значение при определянето на времевата степеня на движение [34]. Ако стойността е много голяма, всички разлики се натрупват в MEI.  $\tau$  има огромен ефект върху времевото представяне на движението.

### 3. Gait Energy Image (GEI)

Използва се процедура за извличане на силуета и се започва с добитите бинарни силуетни последователности. Предварителна обработка включва нормализиране на размера – използва се височината на силуета, за да се определи височина на изображението, последователно хоризонтално подравняване - центриране на горната половина на силуета по отношение на хоризонтален центроид. Фигура 11 показва примери за предварително обработени силуетни кадри в различни последователности, в които човек ходи. Дясното изображение във всеки ред е усредненият силует върху цялата последователност. Както се очаква, изображението на усреднения силует отразява основните форми на човешките силуети и промените им над последователността. Пиксел с по-висока стойност на интензитета означава, че човешкото тяло се появява по-често в тази позиция. Това изображение на усреднения силует се нарича Gait Energy Image (GEI).

Фиг. 11. Примери за нормализирани и подравнени кадри на силует в различни последователности, в които човек ходи. Дясното изображение на всеки ред е изображението на усреднения силует върху цялата последователност Gait Energy Image (GEI)



Получавайки нормализирания размер и хоризонтално подравнената бинарна силуетна последователност, в която човек ходи  $B(x, y, t)$  сиво-ниво GEI  $G(x, y)$  се определя, както следва:

$$G(x, y) = \frac{1}{N} \sum_{t=1}^N B(x, y, t), \quad (9)$$

където  $N$  е броят на кадрите в завършения цикъл на последователността,  $t$  е номера на кадъра в последователността,  $x$  и  $y$  са стойностите на координатите в 2D изображението.

GEI има няколко предимства в сравнение с бинарната силуетна последователност. Като среден шаблон, GEI не е чувствителен към случайни грешки на силуета в отделните кадри. Устойчивостта може допълнително да се подобри, ако премахнем тези пиксели с енергийна стойност по-ниска от прага. Освен това с такъв 2D шаблон, не е нужно да се разделят последователността на силуета в цикли и се изпълнява времево нормализиране по отношение на продължителността на цикъла. Поради това, грешките, които се срещат в тези процедури могат да бъдат избягвани.

В сравнение с бинарна силуетна последователност, загубата на информация от GEI е очевидна. За конкретен пиксел в GEI, знаем само неговата стойност на интензитета, т.е. честотата, с която се появява човешкия силует в тази позиция в продължение на цялата последователност. Въпреки това, би могло частично да се възстанови оригиналната силуетна последователност от GEI според обикновеното човешко ходене. Например, за един пиксел в близост до контура на областта на кракът, GEI стойностите показват, че се

появява силует на това място в 20 от 100 кадъра. Следователно знаем, че 20 кадъра трябва да бъдат тези кадри, където човек ходи, вместо да стои прав, ако не считаме шумът. Може да се разпределят стойностите на GEI към областите крак /ръка на съответните кадри в силуетната последователност. Като цяло, енергийните промени в областта на тялото и главата могат да бъдат считани като шум. Трудно е да се анализира как и в каква степен загубата на информация засяга дискриминиращата силата на GEI като шаблон за индивидуално разпознаване.

#### 4. Structure similarity index measurement (SSIM)

SSIM се използва за намиране на подобие между две (оригиналното и изкривеното) изображения, особено когато има разлика в интензитета, яркостта и контраста. SSIM се основава на три статистически фактора (загуба на корелация, яркостно изкривяване и контрастно изкривяване). Формулата за SSIM е

$$Q = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \cdot \frac{2 \bar{x} \bar{y}}{(\bar{x})^2 + (\bar{y})^2} \cdot \frac{2 \sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}, \quad (10)$$

където първото изображение е  $x = \{x_i | i = 1, 2, \dots, N\}$ , второто изображение е  $y = \{y_i | i = 1, 2, \dots, N\}$ ,  $\bar{x}$  е средноаритметичната стойност на  $x$ ,  $\bar{y}$  е средноаритметичната стойност на  $y$ ,  $\sigma_x^2$  е средноквадратичната грешка на  $x$ ,  $\sigma_y^2$  е средноквадратичната грешка на  $y$ , и  $\sigma_{xy}$  е коефициентът на корелация между  $x$  и  $y$ . Изходът от формула SSIM е качество  $Q$  с динамичен обхват  $[-1, 1]$ . Най-добрата стойност е 1, което означава пълно съвпадение между двете изображения (и двете изображения са еднакви), а -1 означава, че има значителни разлики между двете изображения. За да се изчисли SSIM между две изображения, плъзгащия се прозорецът се движи върху изображението и общата  $Q$  се изчислява като:

$$Q = \frac{1}{M} \sum_{j=1}^M Q_j, \quad (11)$$

където  $M$  е броят на плъзгащите се прозорци върху изображението и  $Q_j$  е индекса на качество на  $j$ -тия прозорец.

### III. Глава

В тази глава ще се разгледа метод за разпознаване на човешки движения, като за целта ще се използват набора от данни на Weizmann. Този набор от данни се записва от неподвижна камера, като по този начин на фоновата сцената също е неподвижна и само обектът се движи. Този набор от данни съдържа 93 видеа за човешка дейност с ниска резолюция (180 x 144, със скорост 50 кадъра в секунда). Наборът от данни съдържа 10 действия: навеждане, скачане с махане на ръце, скачане напред, скачане на място, тичане, скачане на страни, подскачане на един крак, ходене, махане с една ръка и махане с две ръце. Тези действия се извършват от 9 актьора. Всички актьори изпълняват всяко действие веднъж, с изключение на един от актьорите "Лена", които извършват три от действията (тичане, подскачане на един крак и ходене) два пъти. В едно действие, обектът се движи от ляво на дясно, а в друго се движи от дясно на ляво. Наборът от данни на Weizmann се състои от видеа с различни дължини по отношение на броя кадри. Примери с кадри за всяко действие от видео пробите са показани на фигура 12.

Фиг. 12. Примери с кадри за всяко действие (горен ред) навеждане, скачане с махане на ръце, скачане напред, скачане на място, тичане, (долен ред) скачане на страни, подскачане на един крак, ходене, махане с една ръка и махане с две ръце



Предложеният метод се състои от два режима на работа: обучение и тестване. Във всеки режим входното видео преминава през няколко стъпки. Режимът на обучение се използва за обучение на програмата за възможните човешки действия. Режимът на тестване се използва за класифициране и определяне на действието, което се случва във видеото.

#### 1. Режим на обучение

Основната структура на режим на обучение, показан на фиг.13, се състои от пет стъпки.

Фиг. 13. Основна структура на режим на обучение



В началото, след като се прочете видеото за обучение, стъпката за обработката по кадри се стартира.

По време на обработката на кадрите, изваждането на фона се нанася върху всеки кадър от видеото за обучение, за откриване на движещия се обект, където изваденото изображение се получава чрез изваждане на текущия кадър от неговия фон. Всеки пиксел на изваденото изображение показва степента на промяна в движението между кадъра и фона. Чрез праг движещият се обект е открит и изваден. След това извлеченият обект се превръща в двоичен образ, който се състои от 0 и 1 стойности. По време на обработката на кадрите, посока флаг се изчислява, за да се определят действията, които биха могли да се извършат в посока от ляво на дясно или обратно. Предимството на флаг е обединяването на всички действия, които имат движение в една посока (от ляво на дясно). Предимството е, че всички АМІ ще се формират с една единна посока и проблемът с объркването от движенията в различни посоки (например ходене от ляво на дясно или обратно), ще бъде решен. След това двоичното изображение се използва за намиране на медицентър за извлечения обекта, който запълва силуета на обекта. След това се подравнява извлеченото двоично изображение хоризонтално според Y-координата на центъра на тежестта на движението. Подравняването ще образува дискриминативна форма "АМІ", която могат да бъдат използвана като признак за разпознаване.

След обработка на кадри, АМІ се изчислява съгласно уравнения (7), (8) или (9) в зависимост от вида на АМІ.

В последствие обработка АМІ започва с прилагане на допълнителна обработка. Първо, гранична кутия, която е най-малката кутия около получения обект в усредненото подравнено изображение. Второ, граничната кутия на се отрязва. До този момент, всяко видео в тренировъчната серия ще има различно изрязано изображение, в зависимост от

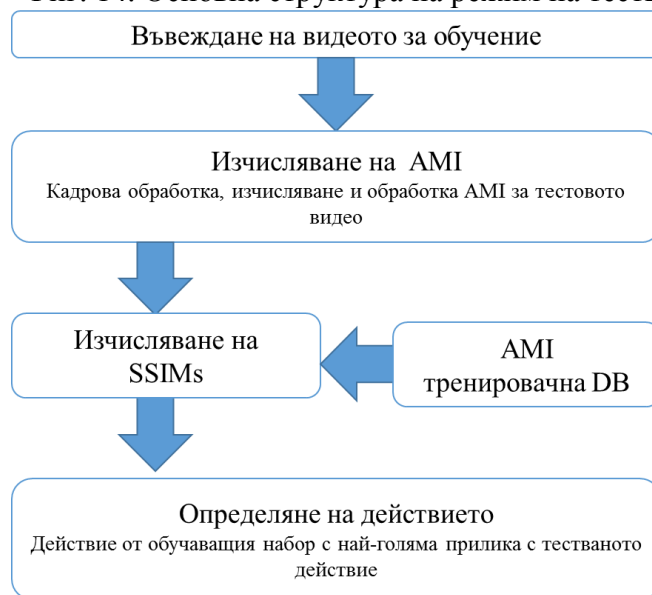
размера (височина и ширина). Тази разлика ще формира изображения с различни размери, които не могат да бъдат използвани директно в SSIM. За да се преодолее този проблем, всички изрязани изображения трябва да бъдат с един и същи размер.

Накрая всички тези обединени изображения (AMIs) се използват като признаци за класифициране и се записват, за да се използват в режима на тестване.

## 2. Режим на тестване

Основната структура за режим тестване, както е показано на Фиг. 14, се състои от няколко стъпки, които са подобни на режим на обучение и други, които са различни.

Фиг. 14. Основна структура на режим на тестване



В началото, всички стъпки в режим на обучение трябва да се направят за видеото за тестване и като резултат AMI се получава за видеото за тестване.

След това, structure similarity index (SSIM) се изчислява между AMI на видеото за тестване и всички записани AMI проби от видеото за обучение. Изчислените индекси, които имат по-високи стойности, имат толкова по-добро сходството, а тези с по-ниски стойности, по-лошо. По този начин, най-високата стойност се взема и действието във видеото за тестване се идентифицира със също действие от видеото на обучение с най-висок SSIM.

SSIM, който се основава на статистическа информация от AMIs изображенията, се използва в режим на тестване като класификатор, чиято цел е да се класифицира действието, което се е случило вътре във видеото за тестване в едно от действията във видеото на обучение.

## IV. Глава

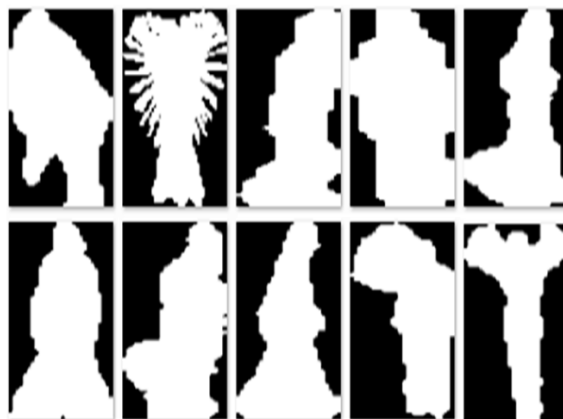
### 1. Експерименти

За тестване на подхода се използват AMEI, AMHI и AGEI, като три различни типа изображения за движение, за алгоритъм. За всеки AMI, се тестват две различни проби (експерименти) в периода от няколко кадъра, използвани от всяка видео проба. В първия експеримент, пълната последователност (всички кадри) на всяка видео проба се използва за образуване на AMIs. Във втория експеримент се използва под последователност (под на първите кадри) във всяка видео проба. Например, започва се само с първите 8 кадъра от всяка видео проба. Най-добрият резултат и номера на кадъра се записват за експеримента под последователност. В експеримента под последователност се забелязва, че в средно 30 кадъра дава най-добър резултат, защото почти всички действия се нуждаят от този брой кадри, за да се завирши периодичността им.

#### 1.1. Подравняване на MEI експерименти

В AMEI експериментите, за режима на обучение, се използва алгоритъма, показан на фиг. 13, за да се обучи програмата за човешките действия. За режим на тестване, се използва алгоритъм на фиг. 14. Изображенията AMEI се изчисляват съгласно уравнението (8), където всяко изображение е двоично (логично) изображение, който се състои само от черни и бели цветове; черен означава, че този пиксел никога не е бил използван във всички кадри на видео пробата, докато бял означава, че пиксел се използва най-малко един път в кадрите. Фиг. 15 показва примери за AMEI изображения, които са получени от 10 различни човешки действия. В фиг. 14 е очевидно, че има известно объркване в сходството между AMEI изображенията (например приликата между бягане и ходене).

Фиг. 15. Примери за (AMEIs) за човешките действия (горен ред) навеждане, скачане с махане на ръце, скачане напред, скачане на място, тичане, (долен ред) скачане на страни, подскачане на един крак, ходене, махане с една ръка и махане с две ръце





Два експеримента са тествани в АМЕІ. Първият експеримент (пълна последователност) се извършва чрез използване на всички кадри във видео пробите. Скоростта на разпознаване на действията достига до 92,47%. Най-добрият резултат за втория експеримент (под последователност) се извършва с помощта само на първите 31 кадъра на видео пробите. Скоростта на разпознаване на действие постигнато в този експеримент е 93,55%. Подробните резултати за тези експерименти са показани в Таблица 1.

Таблица 1. Резултати за разпознаване на човешки движения, използвайки АМЕІ разпознаване.

Човешки действия	АМЕІ разпознаване, използвайки пълна последователност		АМЕІ разпознаване, използвайки под последователност	
	# Точност	% Точност	# Точност	% Точност
Навеждане	9/9	100	9/9	100
Скачан с махане на ръце	9/9	100	9/9	100
Скачане	7/9	77,77	7/9	77,77
Скачане на място	9/9	100	8/9	88,88
Тичане	10/10	100	10/10	100
Скачане настрани	8/9	88,88	8/9	88,88
Подскачане на един крак	7/10	70	8/10	80
Ходене	9/10	90	10/10	100
Махане с една ръка	9/9	100	9/9	100
Махане с две ръце	9/9	100	9/9	100
Крайна точност	86/93	92,47%	87/93	93,55%

В АМЕІ, разпознаването се основава на формата на изображения АМЕІ без никакви промени във вътрешната поява на тези форми, тъй като вътрешната поява на тези форми се състои от една логическа стойност, която е бял цвят. Така, резултатите отразяват някакво объркване поради използването на формата само без вътрешен вид. Тези нежелани скъсявания в резултатите са открити поради приликите на форми между различните действия, особено между действията скачане, подскачане на един крак и тичане, както е показано в матриците на грешките. Таблица 2 показва матрицата на грешките на пълната последователност експеримента и Таблица 3 показва матрицата на грешките на експеримента под последователност.

Таблица 2. Матрица на грешките от AMEI разпознаването за пълна последователност от кадри във видеото.

Човешки действия	Навеждане	Скачане с махане на ръце	Скачане	Скачане на място	Тичане	Скачане на страни	Подскачане на един крак	Ходене	Махане с една ръка	Махане с две ръце	Общ брой движения
Навеждане	9	0	0	0	0	0	0	0	0	0	9
Скачан с махане на ръце	0	9	0	0	0	0	0	0	0	0	9
Скачане	0		7	1	0	0	0	1	0	0	9
Скачане на място	0	0	0	9	0	0	0	0	0	0	9
Тичане	0	0	0	0	10	0	0	0	0	0	10
Скачане на страни	1	0	0	0	0	8	0	0	0	0	9
Подскачане на един крак	0	0	1	0	2	0	7	0	0	0	10
Ходене	0	0	0	0	0	1	0	9	0	0	10
Махане с една ръка	0	0	0	0	0	0	0	0	9	0	9
Махане с две ръце	0	0	0	0	0	0	0	0	0	9	9

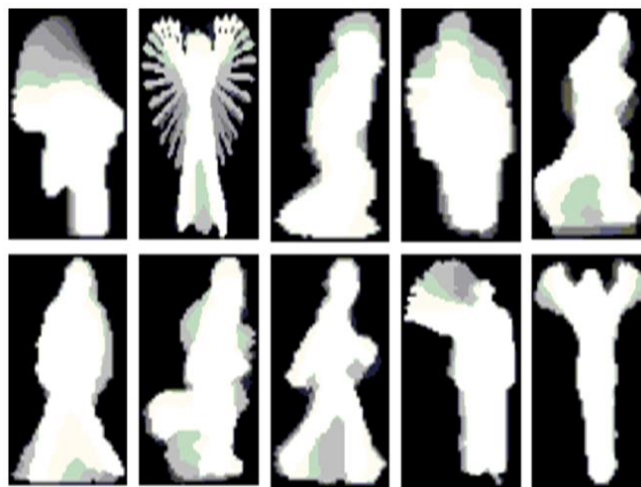
Таблица 3. Матрица на грешките от AMEI разпознаване за най-добра под последователност от кадри във видеото

Човешки действия	Навеждане	Скачане с махане на ръце	Скачане	Скачане на място	Тичане	Скачане на страни	Подскачане на един крак	Ходене	Махане с една ръка	Махане с две ръце	Общ брой движения
Навеждане	9	0	0	0	0	0	0	0	0	0	9
Скачане с махане на ръце	0	9	0	0	0	0	0	0	0	0	9
Скачане	0		7	1	1	0	0	0	0	0	9
Скачане на място	1	0	0	8	0	0	0	0	0	0	9
Тичане	0	0	0	0	10	0	0	0	0	0	10
Скачане на страни	1	0	0	0	0	8	0	0	0	0	9
Подскачане на един крак	0	0	1	0	1	0	8	0	0	0	10
Ходене	0	0	0	0	0	0	0	10	0	0	10
Махане с една ръка	0	0	0	0	0	0	0	0	9	0	9
Махане с две ръце	0	0	0	0	0	0	0	0	0	9	9

### 1.2.Align motion history images (AMHI) експериментите:

В експериментите AMHI, за режим на обучение, използвахме алгоритъма, показан на фиг. 13, за да се обучи програмата за човешките действия. За режим на тестване, се използва алгоритъм на фиг. 14. Техниката „leaving a video out“ се използва в експериментите AMHI. Изображенията AMHI се изчисляват съгласно уравнение (7), където всяко изображение е полутонно изображение, чиито стойности на интензитета са между (0 и 255). Пикселите със стойност 0 означават, че тези пиксела абсолютно не се използват във всички кадри във видео пробата, докато пиксела със стойност  $n$ , означават, че тези пиксели, използвани в  $n$ -ят кадър във видео пробата. По този начин, изображенията AMHI показват историята на движения в кадрите на видео проби, както е показано на фиг. 16.

Фиг. 16. Примери за AMNIs за човешки действия(горен ред) навеждане, скачане с махане на ръце, скачане напред, скачане на място, тичане, (долен ред) скачане на страни,подскачане на един крак, ходене, махане с една ръка и махане с две ръце



Два експеримента са тествани в AMNI. Първият експеримент (пълна последователност) се извършва чрез използване на всички кадри във видео пробите. Скоростта на разпознаване на действие достига 89,25%. Вторият експеримент (под последователност) се извършва чрез използване на различен брой от първите кадри във всички видео проби. Най-добър резултат за втория експеримент се извършва с помощта само на първите 29 кадъра във видео пробите. Скоростта на разпознаване на действие постигнато в този експеримент е 95,70%. Резултатите от тези експерименти са показани в Таблица 4.

Таблица 4. Резултати от пазпизнаване на човешки действия с помоща на АМНІ  
разпознаване

Човешки действия	АМЕІ разпознаване, използвайки пълна последователност		АМЕІ разпознаване, използвайки под последователност	
	# Точност	% Точност	# Точност	% Точност
Навеждане	8/9	88,88	9/9	100
Скачан с махане на ръце	9/9	100	9/9	100
Скачане	6/9	66,66	7/9	77,77
Скачане на място	9/9	100	9/9	100
Тичане	9/10	90	10/10	100
Скачане на страни	8/9	88,88	9/9	100
Подскачане на един крак	6/10	60	8/10	80
Ходене	10/10	100	10/10	100
Махане с една ръка	9/9	100	9/9	100
Махане с две ръце	9/9	100	9/9	100
Крайна точност	83/93	89,25%	89/93	95,70%

В АМНІ, скоростта на разпознаване във втория експеримент (под последователност) е по-добра от тази в първия експеримент (цялата последователност), което се дължи на факта, че по-късните кадрови пикселни стойности ще бъдат изтрети от предишните кадрови пикселни стойности. Този проблем е по вероятно да се появи при експеримента на пълната последователност и по-малко в под последователността. Така объркването в първия експеримент е повече от това във втория.

Разпознаването в АМНІ експериментите се основава и на формата и вътрешния вид на изображения. Въпреки че резултатите отразяват някои разлики когато се сравняват с АМЕ разпознаване, има известно объркване в тези резултати между различните действия, особено сред скацхане и скачане на един крак, както е показано в матрицата на грешките на експеримента на пълната последователност в Таблица 5.

Таблица 5. Матрица на грешките на АМНІ разпознаване за пълна последователност от кадри във видео

Човешки действия	Навеждане	Скачане с махане на ръце	Скачане	Скачане на място	Тичане	Скачане на страни	Подскачане на един крак	Ходене	Махане с една ръка	Махане с две ръце	Общ брой движения
Навеждане	8	0	0	1	0	0	0	0	0	0	9
Скачан с махане на ръце	0	9	0	0	0	0	0	0	0	0	9
Скачане	0		6	2	0	1	0	0	0	0	9
Скачане на място	0	0	0	9	0	0	0	0	0	0	9
Тичане	0	0	0	0	9	0	1	0	0	0	10
Скачане на страни	0	0	0	1	0	8	0	0	0	0	9
Подскачане на един крак	0	0	1	1	2	0	6	0	0	0	10
Ходене	0	0	0	0	0	0	0	10	0	0	10
Махане с една ръка	0	0	0	0	0	0	0	0	9	0	9
Махане с две ръце	0	0	0	0	0	0	0	0	0	9	9

Както е показано на Таблица 6, има някакво объркване между действията: подскачане на един крак и скачане, скачане и скачане на страни и скачане и подскачане на един крак.

Таблица 6. Матрица на грешките на AMNI разпознаване за за най-добра първа под последователност от кадри във видео.

Човешки действия	Навеждане	Скачане с махане на ръце	Скачане	Скачане на място	Тичане	Скачане на страни	Подскачане на един крак	Ходене	Махане с една ръка	Махане с две ръце	Общ брой движения
Навеждане	9	0	0	0	0	0	0	0	0	0	9
Скачан с махане на ръце	0	9	0	0	0	0	0	0	0	0	9
Скачане	0		7	1	0	0	1	0	0	0	9
Скачане на място	0	0	0	9	0	0	0	0	0	0	9
Тичане	0	0	0	0	10	0	0	0	0	0	10
Скачане на страни	0	0	0	0	0	9	0	0	0	0	9
Подскачане на един крак	0	0	2	0		0	8	0	0	0	10
Ходене	0	0	0	0	0	0	0	10	0	0	10
Махане с една ръка	0	0	0	0	0	0	0	0	9	0	9
Махане с две ръце	0	0	0	0	0	0	0	0	0	9	9

### 1.3.Align Gait energy images експерименти:

В експериментите AGEI също се използва технологията „leaves a video out“.

Изображенията AGEI се изчисляват съгласно уравнението (9), където всяко изображение е полутоново, което се състои от реални интензитетни стойности между 0 и 1. Пиксели със стойност 0 са тези, които не се използват във всеки кадър във видео пробата, докато пиксели със стойност 1 са тези, които се използват във всеки кадър и пиксели със стойност 0.5 са тези, които се използват в половината от кадрите, където тези кадри могат да бъдат последователни или непоследователни. С други думи, изображения AGEI показват средните времена на пиксели, които се използват в кадрите на видео пробата, както е показано на фигура 17.

Фиг. 17. Примери за AGEIs за човешки действия (горен ред) навеждане, скачане с махане на ръце, скачане напред, скачане на място, тичане, (долен ред) скачане на страни, подскачане на един крак, ходене, махане с една ръка и махане с две ръце



Again, two experiments are tested in AGEI. The first experiment (complete sequence) is done by using all the frames in the video samples and the action recognition rate achieved was 96.774%. The second experiment (sub sequence) was performed by using different number of the first frames in each video samples. The best result for the second experiment is done by using only the first 33 frames in the video samples. The actions recognition rate that achieved in the second experiment was 98.925% which is the best result that recorded in our approach. The results for both experiments are shown in Table 7.

Отново, два експеримента са тествани в AGEI. Първият експеримент (пълна последователност) се извършва чрез използване на всички кадри във видео пробите и процентът на разпознаване на действия достига 96.774%. Вторият експеримент (под последователност) се извършва чрез използване на различен брой от първите кадри във всяка видео проба. Най-добър резултат за втория експеримент се извършва с помощта само на първите 33 кадра. Процентът на разпознаване на действия, постигнат във втория експеримент е 98,925%, което е най-добрият резултат, Резултатите от двата експеримента са показани в Таблица 7.



Таблица 7. Разпознаване на човешки движения, използвайки AGEI разпознаване

Човешки действия	AMEI разпознаване, използвайки пълна последователност		AMEI разпознаване, използвайки под последователност	
	# Точност	% Точност	# Точност	% Точност
Навеждане	9/9	100	9/9	100
Скачан с махане на ръце	9/9	100	9/9	100
Скачане	9/9	100	9/9	100
Скачане на място	9/9	100	9/9	100
Тичане	10/10	100	10/10	100
Скачане на страни	8/9	88,88	9/9	100
Подскачане на един крак	8/10	80	9/10	90
Ходене	10/10	100	10/10	100
Махане с една ръка	9/9	100	9/9	100
Махане с две ръце	9/9	100	9/9	100
Крайна точност	90/93	96,77%	92/93	98,92%

В AMEI, разпознаването се основава на формата и вътрешния вид на тези форми. В експеримента с използването на всички кадри (пълна последователност) във видео пробата има объркване в три от видеата: две скачане на място със скачане и допълнително объркване между скачане на страни и скачане на място, както е показано в Таблица 8.

Таблица 8. Матрица на грешките на AGEI разпознаването за пълна последователност от кадри във видео.

Човешки действия	Навеждане	Скачане с махане на ръце	Скачане	Скачане на място	Тичане	Скачане на страни	Подскачане на един крак	Ходене	Махане с една ръка	Махане с две ръце	Общ брой движения
Навеждане	9	0	0	0	0	0	0	0	0	0	9
Скачан с махане на ръце	0	9	0	0	0	0	0	0	0	0	9
Скачане	0		9	0	0	0	0	0	0	0	9
Скачане на място	0	0	0	9	0	0	0	0	0	0	9
Тичане	0	0	0	0	10	0	0	0	0	0	10
Скачане на страни	0	0	0	1	0	8	0	0	0	0	9
Подскачане на един крак	0	0	1	0	1	0	8	0	0	0	10
Ходене	0	0	0	0	0	0	0	10	0	0	10
Махане с една ръка	0	0	0	0	0	0	0	0	9	0	9
Махане с две ръце	0	0	0	0	0	0	0	0	0	9	9

В експеримента за под последователност, се използват първите 33 кадъра от всички видео проби. Целта на азпознаването е почти постигната, очаква само една объркана видео проба. Този процент на разпознаване се постига като резултат от използването на форма с оптимален вътрешен вид, който е извлечен от средната стойност на пикселите в ркадрите. Затова метод AGEI позволява формирането на най-добрите разпознаваеми AGEIs и влияят положително на процента на разпознаване. Матрицата на грешките в Таблица 9 показва, че има объркване само във видео пробата между скачане на един крак и тичане в AGEI разпознаването.

Таблица 9. Матрица на грешките на AGEI разпознаване за най-добра първа под последователност от кадри във видео.

Човешки действия	Навеждане	Скачане с махане на ръце	Скачане	Скачане на място	Тичане	Скачане на страни	Подскачане на един крак	Ходене	Махане с една ръка	Махане с две ръце	Общ брой движения
Навеждане	9	0	0	0	0	0	0	0	0	0	9
Скачан с махане на ръце	0	9	0	0	0	0	0	0	0	0	9
Скачане	0		8	0	0	0	0	0	0	0	9
Скачане на място	0	0	0	9	0	0	0	0	0	0	9
Тичане	0	0	0	0	10	0	0	0	0	0	10
Скачане на страни	0	0	0	0	0	9	0	0	0	0	9
Подскачане на един крак	0	0	0	0	1	0	9	0	0	0	10
Ходене	0	0	0	0	0	0	0	10	0	0	10
Махане с една ръка	0	0	0	0	0	0	0	0	9	0	9
Махане с две ръце	0	0	0	0	0	0	0	0	0	9	9

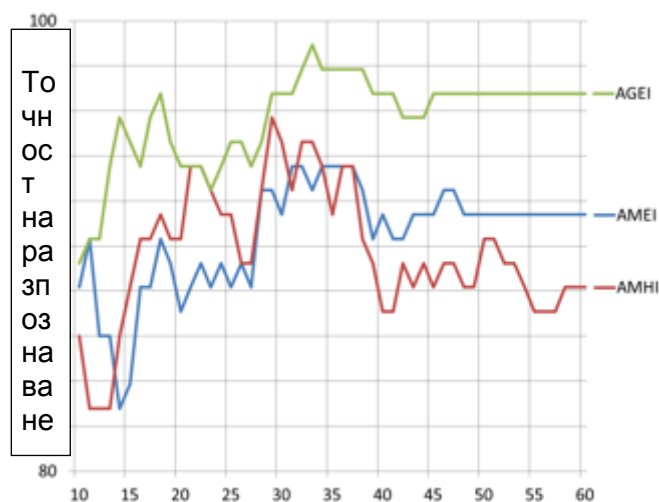
## 2. Съпоставяне на експериментите под последователност и цяла последователност

В първият експеримент (пълна последователност) се използват всички налични кадри във всяка видео проба, за да образуват AMI изображенията, а във втория експеримент (под последователност) се използват определен брой първите кадри във всяко видео. Процентът на разпознаване в експеримента под последователност е по-добър, отколкото експеримента пълна последователност, поради следните факти. Първо, за повечето от действията са нужни около 30 кадъра или по-малко, за да се завърши периода на действие, независимо от началото на действието. На второ място, използването на всички кадри ще се отрази върху резултата лошо, защото по-късните кадри ще се скрият и изтрият предишните кадрови стойности специално във видовете разпознаване AMNI и AMEI.

Във втория експеримент (под последователност) от трите вида на AMIs, се избират първите 31 кадъра за AMEI, 29 кадъра за AMNI и 33 за AGEI. Тези номера са избрани след като експерименталеният тест е приключил за всички кадри на трите вида. Във всеки един от тези експерименти, се започва с избиране на частичен брой от използваните кадри (например първите 10 кадъра, после първите 11 кадъра, и така нататък). Най-добрият процент на разпознаване се постига с броя на използваните кадри и се записва за втория експеримент. Фиг. 18 показва всички резултати за три вида AMI (AMEI, AMNI и AGEI), като се започва от първите 10 кадъра и завършва с първите 60. Ако видео пробата не разполага с достатъчен брой кадри за експеримента на първите

кадри, се използват всички налични кадри. От фиг. 18, е очевидно, че най-добрият процент на разпознаване е за AGEI, когато се използват първите 33 кадъра във всяка видео проба. По същия начин, се забелязва, че най-добрият процент на разпознаване е за AMEI е, когато се използват първите трийсет и един кадъра и най-добрия за AMHI е, когато се използват първите 29 кадъра.

Фиг. 18. Процентни резултати от разпознаването за втория експеримент „под последователност“ за трите типа (AMEI, AMHI, and AGEI).



Номер на използваните кадри

### 3. Сравнение с други алгоритми

Сравнението между резултатите от този и други алгоритми са представени в Таблица 10. Всички други алгоритми също използват данните на Weizmann.

Таблица 10. Сравнение между разглеждания алгоритъм и други, които използват данните на Weizmann.

Методи за разпознаване	Точност
AMEI (пълна последователност)	92,47%
AMEI (под последователност)	93,55%
AMHI (пълна последователност)	89,25%
AMHI (под последователност)	95,70%
AGEI (пълна последователност)	96,77%
AGEI (под последователност)	98,93%
(Gorelick et al., 2007)	100,00%
(Ali and Shah, 2010)	95,75%
(Naiel et al., 2011)	94,38%
(Yuan et al., 2009)	92,90%
(Yang et al., 2009)	92,80%
(Niebles and Fei-Fei, 2007)	72,80%

Процентът на разпознаване с помощта на AMIs показва много добри резултати в период от процентна точност, особено в експериментите под последователност във всички три типа (AMEI, AMHI и AGEI), както е показано в таблица 10. Също така, в резултатът на AGEI в двата експеримента (пълна и под последователност) е по-добър в сравнение с другите два вида (AMEI and AMHI), защото решават проблема на изтрети пикселни стойности в по-ранните кадри от пикселните стойности в по-късните кадри. Най-добрият резултат е постигнат в експеримента под последователност на AGEI, както се наблюдава ясно в Таблица 10. Третият най-добър резултат е за експеримента пълна последователност на AGEI.

## V. Глава

В тази глава ще се представи икономическа оценка на резултатите и техническа ефективност. За целта е представена таблица 11.

Таблица 11. Икономическа оценка

	Цена	Чсове	
Matlab for student	245 лв.		
Lenovo G 710	1000 лв.		
Човешки труд	15 лв/ч	15 часа	Общо 1470 лв

## VI. Глава

В тази глава ще се разгледа високото ниво от фиг.1 от първа глава, т.е. приложенията за разпознаване на движения. Както е показано на фигурата главните сфери са три – за наблюдение, за забавление и здравеопазване.

### 1. Среди за наблюдение

Приложенията за разпознаване на човешката дейност в системи за наблюдение е фокусирано главно върху автоматичното проследяване на лица и тълпи, така че да се поддържа персонал по сигурността, за да наблюдава и разбира дейността, в резултат на разпознаване на престъпника и откриване на подозрителни дейности..

Повечето охранителни системи за наблюдение са оборудвани с няколко камери и изискват човек, който да наблюдава екрани. Чрез прилагането на автоматични техники за разпознаване на човешка дейност във видео-базирани охранителни системи, може ефективно да се намали натовареността на персонала по сигурността, тъй както систематично се сигнализира, когато са открити нарушения в сигурността, с цел предотвратяване на потенциално опасни ситуации. Някои типични сценарии са, както следва:

Откриването и проследяването на хора е една от първите цели на охранителната система за наблюдение. Nakazawa et al. [35] предлага метод за проследяване на човека използвайки свързан в мрежа визуални системи, всяка от която се състои от камера и модул за обработка на изображения. Всяка визуална система има две задачи, задачата за проследяване и задачата на придобиване. Ако има хора във видимата област на визуалната система, системата за проследяване проследява човешката позиция и предава резултатите на другите системи. В случай че визуалната системата не засича човек, след това системата за придобиване трябва да намери лицето в своето изображение за да бъде проследено. Bodor et al. [36] използват Калман филтри за проследяване на местоположението и изминатия път на всеки пешеходец в области с висок пешеходен трафик, а след това резултатите от проследяването се използва за по-нататъчна обработка за откриване на съмнителни поведения, като например за въвеждане на "защитена зона", тичането или движението хаотично, движещи се хора срещу движението, открадване на чанти или други предмети.

Навъртането е подозрително поведение, което привлича много изследователи, което се дължи на факта, че доста често води до необичайни ситуации, като например продаване на наркотици, банков обир, джебчийство и т.н. Bird et al. [37] представя визуално базиран метод за откриване на лица дебнещи на автобусни спирки. С помощта на стационарна камера за наблюдение на автобусната спирка, системата взема снимки на лица и след това ги използва, за да ги класифицира използвайки метод за разпознаване по външен вид. Характеристиките, използвани за да се съпоставят лица се основават само на цветовете на дрехите. За да се определи дали дадено лице се навъртало на същата автобусна спирка, вместо да се използва човешка техника за проследяване от видео

записи, чрез снимките в базата данни може да се използват за да се прецени колко дълго едно лице е било налице.

## 2. Среди за забавление

Разпознаването на човешка дейност може да бъде използвана за разпознаване на спорт, танци и игри.

Разпознаването на спорт е ефективно признато чрез използване DBNs, които могат да генерират йерархична описание за видео събития, включително боулинг, ски спускане, голф суинг и ски скокове, записани от реални сценарии, с затрупана фон и движещи се камери.

Една от най-популярните развлекателни дейности е игране на видео игри. Редица методи са разработени за тази цел. Ричард и др. [38], разработена Pfnder като система в реално време за проследяване на хора и тълкуване техните действия чрез използване на мулти-клас статистически модел, основан на цвят и форма, за да се получат по главата, ръцете и краката позиции в различни условия на гледане. Pfnder се използва успешно в няколко различни приложения човешки интерфейс, например, за да се движите на 3D виртуална игра, или да пусне играча на определено място във виртуална стая, която е населена с виртуални обитатели от реално време 3D компютърни графики на базата на видео на живо. В [39], Нuo et al. представя метод за човешки улавяне на движение и разпознаване. Човешкия торс и ръцете са сегментирани от цялото тяло и се проследяват с течение на времето. А 2D модел се използва за откриване и проследяване на торса, докато моделът за цвета на кожата се използва за проследяване на ръце.

## 3. Здравеопазване

Приложенията за разпознаване на дейност в здравните системи анализират и разбират дейностите на пациентите, така че да се улесни здравни работници за диагностика, лечение и грижа за пациентите, което води до подобряване на надеждността на диагностиката, намаляване на работното натоварване за медицинския персонал, съкращаване на болничния престой на пациентите.

### 3.1. Наблюдение на ежедневните действия

Наблюдение на ежедневните действия се съсредоточава главно върху обучението и разпознаването на ежедневните житейски дейности на възрастните хора у дома. Предлагащите системи са, за да предоставят на възрастните възможност да живеят в безопасност, независимо и удобно. За да се постигне това, повечето системи непрекъснато улавят движенията на възрастните у дома, автоматично разпознават техните дейности, както и откриват постепенни промени в базовите дейности като мобилност, функционални увреждания, психични проблеми, както и спешните предупредителни знаци на необичайни дейности, като например падане или получаване на инсулт.

### 3.2. Приложения за рехабилитация

Традиционните рехабилитационни системи често изискват пациентите да се подложат на няколко клинични посещения за упражненията (терапия), докато тяхното



здраве не се възстанови напълно. Такива клинични посещения могат да бъдат избегнати чрез използването на иновативни рехабилитационни системи, които се инсталират в домашни условия и пациента сам се грижи за себе си с помощта на видео-базирани техники за разпознаване на дейност. Освен това, чрез непрекъснато наблюдение на ежедневните дейности и ходовете, ранните симптоми на някои болести могат да бъдат своевременно открити, така че диагнозата и интервенции са по-подходящи.

## VII. Глава

В тази дипломна работа беше представено разпознаването и определянето на човешки действия, с помощта на видео пробите на Weizmann. За да се постигне това, бяха представени човешките действия в aligned motion images (AMIs) и използвани тези изображения като признак за разпознаване на действията. За класификацията се използва structure similarity index measure (SSIM), който постигна много добри резултати в периода на разпознаване с голям процент точност, както е показано в експерименталните резултати и в сравнението направено в таблица 10 от глава 4.

Три вида AMI изображения (AMEI, AMHI и AGEI) са използвани в експериментите. Във всеки вид, два различни експеримента (пълна последователност и под последователност) се провеждат в границата на броя използвани кадри за сформирането на AMI. Като цяло, резултатите от експериментите с под последователност са по-добри, отколкото тези с пълната последователност, защото повечето действия са били завършени в тези под последователни кадри, свели до минимум смущението между AMI изображенията. Измежду трите вида AMI, резултатите от вида AGEI са по-добри от резултатите на другите, защото при него е решен проблема с изтрети пиксели в по-ранните кадри.

В заключение, експериментът на AGEI използва под последователни кадри във всички видео проби, постигайки 98,93%, което е най-точното разпознаване постигнато във всички експерименти на алгоритъма. Освен това, AGEI не разрешава въпроса за използваните по-късно пиксели, обхващащи по-ранни такива.

За подобряване на резултатите в бъдеще може да се смесят трите метода в един по-ефективен метод, да се използва друга база данни (видео), да се промени фона (от неподвижен към подвижен), да се включи взаимодействието на хората с предмети, животни и други хора, както и препятствия на преден план във видеото.

## Използвана Литература

1. Wren, C.R.; Azarbayejani, A.; Darrell, T.; Pentland, A.P. Pfunder: Real-time tracking of the human body. *IEEE Trans. Pattern Anal. Mach. Intell.* 1997, 19, 780–785.
2. Cucchiara, R.; Grana, C.; Piccardi, M.; Prati, A. Detecting moving objects, ghosts, and shadows in video streams. *IEEE Trans. Pattern Anal. Mach. Intell.* 2003, 25, 1337–1342.
3. Seki, M.; Fujiwara, H.; Sumi, K. A Robust Background Subtraction Method for Changing Background. In *Proceedings of Fifth IEEE Workshop on Applications of Computer Vision*, Palm Springs, CA, USA, 4–6 December 2000; pp. 207–213.
4. Blank, M.; Gorelick, L.; Shechtman, E.; Irani, M.; Basri, R. Actions as Space-time Shapes. In *Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV)*, Beijing, China, 17–21 October 2005; Volume 2, pp. 1395–1402.
5. Kumari, S.; Mitra, S.K. Human Action Recognition Using DFT. In *Proceedings of the third IEEE National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, Hubli, India, 15–17 December 2011; pp. 239–242.
6. Lowe, D.G. Object Recognition from Local Scale-invariant Features. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, Kerkyra, Greece, 20–25 September 1999; Volume 2, pp. 1150–1157.
7. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, 60, 91–110.
8. Dalal, N.; Triggs, B. Histograms of Oriented Gradients for Human Detection. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, San Diego, CA, USA, 20–26 June 2005; Volume 1, pp. 886–893.
9. Huo, F.; Hendriks, E.; Paclik, P.; Oomes, A.H.J. Markerless Human Motion Capture and Pose Recognition. In *Proceedings of the 10th IEEE Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, London, UK, 6–8 May 2009; pp. 13–16.
10. Veeraraghavan, A.; Roy-Chowdhury, A.K.; Chellappa, R. Matching shape sequences in video with applications in human movement analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, 27, 1896–1909.
11. Sempena, S.; Maulidevi, N.U.; Aryan P.R. Human Action Recognition Using Dynamic Time Warping. In *IEEE International Conference on Electrical Engineering and Informatics (ICEEI)*, Bandung, Indonesia, 17–19 July 2011; pp. 1–5.
12. Duong, T.V.; Bui, H.H.; Phung, D.Q.; Venkatesh, S. Activity Recognition and Abnormality Detection with the Switching Hidden Semi-Markov Model. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 838–845.
13. Luo, Y.; Wu, T.; Hwang, J. Object-based analysis and interpretation of human motion in sports video sequences by dynamic Bayesian networks. *Comput. Vis. Image Underst.* 2003, 92, 196–216.

14. Schuldt, C.; Laptev, I.; Caputo, B. Recognizing Human Actions: A Local SVM Approach. In Proceedings of the 17th IEEE International Conference on Pattern Recognition (ICPR), Cambridge, UK, 23–26 August 2004; Volume 3, pp. 32–36
15. Agarwal, A.; Triggs, B. Recovering 3D human pose from monocular images. *IEEE Trans. Pattern Anal. Mach. Intell.* 2006, 28, 44–58.
16. Foroughi, H.; Naseri, A.; Saberi, A.; Yazdi, H.S. An Eigenspace-based Approach for Human Fall Detection Using Integrated Time Motion Image and Neural Network. In Proceedings of IEEE 9th International Conference on Signal Processing (ICSP), Beijing, China, 26–29 October 2008; pp. 1499–1503
17. Bodor, R.; Jackson, B.; Papanikolopoulos, N. Vision-based Human Tracking and Activity Recognition. In Proceedings of the 11th Mediterranean Conference on Control and Automation, Rhodes, Greece, 18–20 June 2003; Volume 1, pp. 18–20.
18. Ribeiro, P.C.; Santos-Victor, J. Human Activity Recognition from Video: Modeling, Feature Selection and Classification Architecture. In Proceedings of the International Workshop on Human Activity Recognition and Modelling (HAREM), Oxford, UK, 9 September 2005; Volume 1, pp. 61–70.
19. Ben-Arie, J.; Wang, Z.; Pandit, P.; Rajaram, S. Human activity recognition using multidimensional indexing. *IEEE Trans. Pattern Anal. Mach. Intell.* 2002, 24, 1091–1104.
20. Lucas, B.D.; Kanade, T. An Iterative Image Registration Technique with An Application to Stereo Vision. In Proceedings of the 7th International Joint Conference on Artificial Intelligence, Vancouver, B.C., Canada, 24–28 August 1981.
21. Lin, C.; Hsu, F.; Lin, W. Recognizing human actions using NWFE-based histogram vectors. *EURASIP J. Adv. Signal Process.* 2010, 2010, 9.
22. Sedai, S.; Bennamoun, M.; Huynh, D. Context-based Appearance Descriptor for 3D Human Pose Estimation from Monocular Images. In Proceedings of IEEE Digital Image Computing: Techniques and Applications (DICTA), Melbourne, VIC, Australia, 1–3 December 2009; pp. 484–491
23. Rabiner, L.; Juang, B. *Fundamentals of Speech Recognition*; Prentice Hall: Englewood Cliffs, NJ, USA, 1993.
24. Rabiner, L.; Juang, B. An introduction to hidden Markov models. *IEEE ASSP Mag.* 1986, 3, 4–16.
25. Tipping, M.E. Sparse Bayesian learning and the relevance vector machine. *J. Mach. Learn. Res.* 2001, 1, 211–244.
26. Jain, A.K.; Duin, R.P.W.; Mao, J. Statistical pattern recognition: A review. *IEEE Trans. Pattern Anal. Mach. Intell.* 2000, 22, 4–37.
27. Liu J, Zhang N (2007) Gait history image: a novel temporal template for gait recognition. *IEEE international conference Multimedia and Expo*, pp 663–666
28. Bobick A, Davis J (1996) An appearance-based representation of action. *International conference on, pattern recognition*, pp 307–312

29. Bobick A, Davis J (2001) The recognition of human movement using temporal templates. *IEEE Trans Pattern Anal Mach Intell* 23(3):257–267
30. Meng H, Pears N, Bailey C (2006) Human action classification using SVM\_2 K classifier on motion features. *Multimedia content representation, classification and security. LNCS* 4105:458–465
31. Davis J (1996) Appearance-based motion recognition of human actions. M.I.T. Media lab perceptual computing group technical report 387
32. Alahari K, Jawahar C (2006) Discriminative actions for recognizing events. *Indian conference on computer vision, graphics and image processing*, pp 552–563
33. Meng H, Pears N, Bailey C (2006) Human action classification using SVM\_2 K classifier on motion features. *Multimedia content representation, classification and security. LNCS* 4105:458–465
34. Bobick A, Davis J (2001) The recognition of human movement using temporal templates. *IEEE Trans Pattern Anal Mach Intell* 23(3):257–267
35. Nakazawa, A.; Kato, H.; Inokuchi, S. Human Tracking Using Distributed Vision Systems. In *Proceedings of IEEE Fourteenth International Conference on Pattern Recognition, Brisbane, Qld., Australia, 20 August 1998; Volume 1*, pp. 593–596.
36. Bodor, R.; Jackson, B.; Papanikolopoulos, N. Vision-based Human Tracking and Activity Recognition. In *Proceedings of the 11th Mediterranean Conference on Control and Automation, Rhodes, Greece, 18–20 June 2003; Volume 1*, pp. 18–20.
37. Bird, N.D.; Masoud, O.; Papanikolopoulos, N.P.; Isaacs, A. Detection of loitering individuals in public transportation areas. *IEEE Trans. Intell. Transp. Syst.* 2005, 6, 167–177.
38. Wren, C.R.; Azarbayejani, A.; Darrell, T.; Pentland, A.P. Pfunder: Real-time tracking of the human body. *IEEE Trans. Pattern Anal. Mach. Intell.* **1997**, 19, 780–785.
39. Huo, F.; Hendriks, E.; Paclik, P.; Oomes, A.H.J. Markerless Human Motion Capture and Pose Recognition. In *Proceedings of the 10th IEEE Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), London, UK, 6–8 May 2009; pp. 13–16.*

Списък на използваните означения:

- STV –пространствено-времеви обем
- DFT – Фурие трансформация
- SIFT – scale-ivariant feature transform
- HOG – Хистограма на ориентираните градиенти
- HMM – Скрит Марков модел
- DBN – Динамична Байсова мрежа
- SVM – Машина за поддържащи вектори
- RVM – Relevant Vector Machine
- ANN – Изкуствена невронна мрежа
- K-NN – К-близки съсед
- GMM – Смесен Гаусов модел
- LKT – Lucas-Kanade-Tomasi
- DTW – Dinamic time warping
- MHI – Motion History Image
- MEI – Motion Energy Image
- SSIM – Structure Similarity Index Measurment
- GEI – Gait Energy Image
- AMI – Aligned Motiom Image
- AMHI – Aligned Motion History Image
- AMEI – Aligned Motion Energy Image
- AGEI – Aligned Gait Energy Image

## Съдържание:

1. Увод	стр.2
2. Глава 1 – Състояние на проблема по литературни данни	стр.3
3. Глава 2 – Теоретично решение на поставената задача	стр.13
4. Глава 3 – Описание на апаратната и софтуерната част	стр.21
5. Глава 4 – Изчислителна част	стр.24
6. Глава 5 – Икономическа оценка	стр.38
7. Глава 6 – Приложимост на дипломната работа	стр.39
8. Глава 7 – Изводи и претенции за самостоятелно получени резултати	стр.42
9. Използвана литература	стр.43
10. Списък на използваните означения	стр.46