Treball fi de carrera

ENGINYERIA TÈCNICA EN INFORMÀTICA DE SISTEMES

Facultat de Matemàtiques
Universitat de Barcelona

Programació Dinàmica per al Reconeixement de Gestos:

Aplicació a una nova base de dades de Subjectes, Extremitats, Accions i Interaccions

Jordi Suñé Fontanals

Directors: Sergio Escalera Guerrero
Miguel Ángel Bautista Martín
Realitzat a: Departament de Matemàtica Aplicada i Anàlisi. UB

Barcelona, 20 de gener de 2012

Agraïments

Després de molta feina i temps invertit en aquest projecte, és una tasca complicada la d'agrair a tota la gent que amb la seva presència durant aquest període m'ha ajudat ha arribar a aquest moment. M'agradaria per tant dedicar-lo a totes elles, que són entre d'altres:

Al Dr. Sergio Escalera i a Miguel Angel Bautista, els meus dos tutors, per la seva dedicació, esforç i sobretot paciència mentre m'han transmès els coneixements que han sigut fonamentals per la finalització del projecte. En tot moment han estat disposats a trobar una estona de temps per poder parlar sobre els problemes i dubtes que anaven sortint durant l'elaboració del projecte, per trobar-ne una solució i poder-lo finalitzar.

A Griselda Ginestà Priego per animar-me a continuar endavant en tot moment, no només en l'elaboració del projecte sinó durant tots els anys que he estat cursant la carrera.

A tota la meva família, en especial als meus pares i la meva germana que durant tota la meva vida han estat allà oferint-me ànims, comprensió, afecte i suport.

A tots els companys que m'han acompanyat durant aquesta etapa de la meva vida: Eduard Rafael, Daniel Vidal, Miguel Muñoz, Adrià Monguillot, Raul Ruiz, David Trigo i Helena Orihuela. I també a Jordi Serral, Ivan Aibar, Antoni Rocafort (Roka tio), Carles Riera, Francisco Villalba (Paco), Albert González, Pau Almirall, Adrián Hermoso, Guillem Espejo, Dani Sánchez, Oscar Gamiz, Marc Serra, Ernest Pastor i Manel García amb els que he compartit molts bons moments.

Als meus amics: Jordi Masip, Guillem Villena, Enric Martret, Pablo Montesinos, Xavier Espí, Elisabet Fuentes, Albert Ginés, Carlos Cuadrado, Pedro de la Calle, Miguel Angel Arcos, Raúl Morado, Jesús Galván, Eloy Fernández i Victor de Blas.

I per últim com no mencionar al meu company de projecte durant l'elaboració de la base de dades, Tomás Pérez, ja que entre els dos hem aconseguit generar una base de dades que esperem que sigui d'utilitat en projectes futurs.

Sense vosaltres no hagués estat possible, moltíssimes gràcies.

Resum

Aquest projecte implementa i analitza l'eficàcia de l'algorisme "Dynamic Time Warping (DTW)" per el reconeixement d'accions. S'han utilitzat quatre tipus de descriptors amb característiques diferents per validar el mètode amb major taxa de reconeixement. Totes les dades emprades en el projecte han estat extretes manualment a partir de 10 mostres de vídeos on hi apareixen accions pre-definides. Per la implementació de l'algorisme, el llenguatge de programació que s'ha usat és Matlab, el mateix que s'ha utilitzat per crear l'aplicació per extreure les mostres dels vídeos.

Abstract

This project implements and analyzes the effectiveness of the algorithm "Dynamic Time Warping (DTW)" for recognizing gestures. It has been using 4 different feature descriptors to validate the method with the highest recognition rate. All data used in the project has been manually extracted from 10 videos samples where there are pre-defined actions. For the algorithm implementation the programing language used is Matlab, which also has been used to create the aplication to extract the samples from the vídeos.

Resumen

Este proyecto implementa y analiza la eficacia del algoritmo "Dynamic Time Warping(DTW)" para el reconocimiento de acciones. Se han utilizado cuatro tipos de descriptores con características diferentes para validar el método con mayor porcentaje de reconocimiento. Todos los datos usados en el proyecto han estado extraidos manualment a partir de 10 vídeos donde aparecen acciones predefinidas. Para la implementación del algoritmo, el lenguaje de programación que se ha usado es Matlab, el mismo que se ha utilizado para crear la aplicación para extraer las muestras de los vídeos.

Índex

1. Introducció

1.1. Context del problema

Per la realització d'aquest projecte s'han de definir tres conceptes:

- Què és una acció i quins tipus d'accions podem reconèixer.

- Com s'obtenen les mostres de la base de dades.

- Quina informació necessitem conèixer de les mostres de la base de dades.

1.1.1. Definició d'acció

Una acció és el comportament que du a terme un individu enfront l'entorn on té en compte l'existència d'altres individus que interaccionen amb el mateix medi. Aquest comportament ha de ser fàcilment detectable. Així doncs, nosaltres tindrem en compte les diferents accions dividides en dos subgrups:

Accions realitzades per un sol individu:

- Saludar.
- Senyalar.
- Aplaudir.
- Saltar.
- Ajupir-se
- Caminar.
- Córrer.

Accions en les que necessitem la interacció com a mínim entre dos individus:

- Donar la mà.
- Abraçar.
- Donar-se dos petons.
- Barallar-se.

Si ens fixem podem observar que són accions fàcilment reconeixibles, i bastant diferents les unes de les altres. Aquestes han estat escollides pel motiu que totes les persones les realitzen més o menys de la mateixa manera. D'entre totes cal comentar que l'acció de barallar-se és la que conté moviments més aleatoris ja que cada individu la pot realitzar de dues maneres diferents: atacant o defensant, a més a més, les extremitats usades en l'acció atacant també seran diferents (des de cops de puny fins a puntades de peu). Aleshores probablement serà l'acció que més difícil ens serà d'identificar.

Per tal d'identificar una acció ens centrarem en el moviment de les parts del cos que hem decidit prendre com a referència, cobrint en tot moment la totalitat de cos de l'individu que apareix en el vídeo.

1.2. Motivació

Al veure que no es disposaven de gaires mostres per a realitzar un estudi sobre el reconeixement de patrons d'accions humanes es va decidir crear-ne una amb una gran quantitat de mostres, no només per la realització de dos projectes de final de carrera, sinó que existeix l'esperança que donat l'esforç que ha comportat generar-la pugui ser utilitzada en un futur per altres persones, ja que actualment no existeix una base de dades d'aquest nivell de volum de dades, etiquetada fotograma a fotograma, a nivell d'investigació.

També es vol veure quin dels quatre sistemes de càlcul de descriptors dels que em plantejat és millor alhora de reconèixer, mitjançat l'algorisme Dynamic Time Warping[1][2][3] (DTW), patrons d'accions. Així doncs, ens centrarem en veure l'efectivitat de l'algorisme i finalment traurem unes conclusions que inclouran les seves avantatges, els inconvenients i que podríem fer per millorar-lo.

1.3. Estat del l'art

El reconeixement de gestos és una aplicació de l'àrea de visió per computador en la que un conjunt de tècniques de processament d'imatges i anàlisis de sèries temporals són emprades per a fer que l'ordinador entengui un gest capturat per una càmera. Després del processat d'aquestes imatges el que s'utilitza són algorismes per reconèixer els patrons d'acció mitjançant l'anàlisi de gestos.

Per nombrar algun altre exemple d'aquests algorismes a part del Dynamic Time Warping, que és l'emprat en aquest projecte, comentarem breument el Model Ocult de Markov (HMM)[4] [5].

L'objectiu d'aquest és també el de determinar els paràmetres desconeguts en una seqüència a partir d'uns paràmetres observables, però la diferència radica en que un és un model paramètric amb una estructura fixe i un gran nombre de paràmetres a estimar estadísticament i l'altre és no paramètric, és a dir, a partir d'una seqüència no coneguda, es compara la distància amb els patrons emmagatzemats prèviament. Per concloure, m'agradaria citar que segons un estudi[6] , podem dir que amb grans quantitats de mostres el Dynamic Time Warping és lleugerament superior al Model Ocult de Markov.

1.3.1. Aplicacions actuals

Actualment les àrees que més beneficis estan tenint amb el reconeixement d'accions són: el reconeixement de llenguatge de signes, la robòtica, la telemedicina, la realitat virtual, etc. En resum, aplicacions per a millorar la interacció i l'enteniment entre l'home i les màquines.

En el present ja disposem de dispositius que són capaços de captar les accions en un entorn en tres dimensions (amb profunditat) per a les nostres activitats de lleure, com per exemple el dispositiu Kinect de Microsoft.[7]

1.4. Resum de la proposta

Tenim dos objectius principals en aquest projecte. El primer serà crear una base de dades amb les mostres suficients de les accions especificades (on participaran diferents individus) i el segon verificar quin dels mètodes d'obtenció de descriptors, utilitzats en l'algorisme Dynamic Time Warping és el més idoni per tal d'aconseguir una major taxa d'encert en el reconeixement d'accions.

Per tal d'aconseguir aquests objectius existeixen aquest requeriments:

- Gravar 10 vídeos on apareixeran diversos individus realitzant les diferents accions.
- Dividir els vídeos en fotogrames per procedir a l'etiquetatge de les accions, que descriurem posteriorment (secció 4.1).
- Crear una aplicació que ens servirà per etiquetar cada part del cos i anar emmagatzemant cada mostra en la base de dades.
- Crear l'algorisme Dynamic Time Warping (descrit en l'apartat 2.1.1) per tal de calcular la similitud entre una acció envers una seqüència sencera d'accions i procedir a la seva identificació.
- Crear un algorisme que ens calculi el percentatge d'encert de cadascun del mètodes emprats.

1.5. Organització de la memòria

La memòria està estructurada en 5 blocs ben diferenciats:

- Primer bloc:
  Aquest correspon a la part on ens trobem actualment, on es farà una breu explicació del contingut de la memòria i alhora mostrarà una visió general del contingut del projecte.

- Segon bloc:
  En aquest apartat es comentaran, els mètodes emprats per la identificació d'accions, els requeriments de l'aplicació (amb el seu corresponent

diagrama de casos d'ús), la planificació del projecte i per últim els costos econòmics de l'aplicació. Així doncs en aquest apartat farem una abstracció del que volem fer.

- Tercer bloc:

  En aquesta secció profunditzarem els detalls de l'algorisme donat que ens centrarem en el seu disseny i la seva implementació, explicant les tecnologies de programació emprades, comentant les parts importants del codi i analitzarem el problemes sorgits durant la programació i les seves solucions.

- Quart bloc:

  Aquí descriurem com s'han obtingut les mostres que formen la base de dades, els punts tinguts en compte alhora de donar com a bona una mostra, el procediment d'etiquetatge i el perquè dels formats escollits alhora de guardar les dades.

- Cinquè bloc:

  En el cinquè bloc es mostraran els resultats en quatre taules (una per cada mètode emprat) i es comentaran els valors obtinguts, el protocol de validació que hem seguit i amb quines dades em tractat de la base de dades per obtenir-ne els resultats.

- Sisè bloc:

  Per últim mostrarem les referències i conclusions extretes durant l'elaboració d'aquest projecte.

2. Anàlisi

2.1. Mètodes emprats

En primer lloc hem obtingut per cada vídeo 4 matrius de distàncies que contenen uns valors numèrics segons la fórmula que hem emprat alhora de fer el seu càlcul, aquestes matrius a partir d'ara seran anomenades matrius de descriptors.

2.1.1. Dynamic Time Warping per al reconeixement gestual amb detecció de començament i final de seqüència

Per la realització dels experiments s'ha emprat l'algoritme Dynamic Time Warping per fer Template matching, aquest ens permet veure el grau de similitud entre dues seqüències en el temps, tenint en compte que la seva duració pot ser variable. Per tal de poder aplicar aquest algoritme necessitarem tres paràmetres:

- Una seqüència en la que intentarem trobar un patró. Aquesta serà la que contindrà totes o algunes de les accions realitzades per un altre individu que no sigui el mateix del que provenen les accions que estem buscant.

- La seqüència que contindrà el patró que estem buscant, és a dir, l'acció en concret que nosaltres estem buscant en el nostre experiment.

- Un valor numèric que serà el que ens doni un percentatge d'encert més òptim (explicació en el punt 5.1.1), que ens servirà com valor de tall i que serà a partir del qual acceptem la resta de paràmetres com a similars.

La matriu de distàncies que es generarà a partir d'aquest càlcul ha de tenir tantes columnes com el nombre de fotogrames de la seqüència d'entrada i tantes files com vectors de dades tingui.

Ara que ja sabem les dimensions, tenim que procedir a omplir-la dels valors que ens serviran per identificar patrons. Aquest càlcul es realitza de la següent forma:

- Calculem la distància euclidiana[8] entra cada parell de dades (fila i columna).

- Al valor obtingut li tenim que sumar el mínim dels tres valors que l'envolten. Tal com podem veure en la següent taula:

Valor superior esquerra	Valor superior
Valor esquerra	Valor = valor distància + Mínim(Valor esquerra, Valor superior esquerra, Valor superior)

Taula 1. Exemple d'un valor a la matriu de distàncies

En el moment que tenim tota la matriu emplenada queda veure com decidim prendre un valor com a bo, en aquest instant entra en acció el valor numèric que ens dona millors resultats, anomenat valor de tall. Amb aquest el que farem serà comparar els valors de l'ultima fila i en cas de trobar-ne un de igual o més petit és podria treure la conclusió que hem reconegut una acció semblat a la del patró que nosaltres hem volgut comparar. Una vegada tenim identificat aquest valor hem de trobar el seu punt d'origen per lo que tindrem que intentar refer el camí.

Per la detecció del camí d'inici a fi s'ha decidit emplenar la primera línia de la matriu de distàncies amb zeros, donat que en el moment de realitzar la comparació, el valor més baix que podem obtenir dins de la matriu és el 0, així doncs quan hem seguit el camí i hem arribat al valor zero, significarà que hem recorregut el trajecte des de la fila inferior on hem localitzat el valor de tall, fins la primera. La posició del zero en aquesta fila ens coincidirà amb el número de fotograma on s'iniciarà la possible acció que hem detectat. I ja tindrem les dues dades necessàries el fotograma d'inici i el fotograma final de l'acció reconeguda.

Per tal de veure el camí, el que farem serà agafar el valor identificat i anar buscant el mínim entre tres valors:

- Valor superior.
- Valor esquerra.
- Valor superior esquerra.

En la taula següent es mostra el procediment que segueix l'algorisme una vegada ha trobat el valor semblant al que nosaltres hem acotat, en aquest cas el valor és igual a 9.5.

	Dada 1	Dada 2	Dada 3	Dada 4	Dada 5	Dada 6
	0	0	0	0	0	0
Dada acció 1	1.1023	2.977	3.0002	6.3425	9.5436	13.2245
Dada acció 2	4.7645	4.2345	3.567	5.1231	11.1234	14.1325
Dada acció 3	3.4345	8.3275	4.7563	4.076	10.5543	5.2345
Dada acció 4	7.2346	9.8742	6.1298	9.3424	8.032	12.4563
Dada acció 5	13.7654	14.9564	10.0345	12.3476	9.562	9.384

Taula 2. Exemple del recorregut del camí del fotograma final a l'inicial

A continuació es mostra una matriu de distàncies real amb el camí marcat:

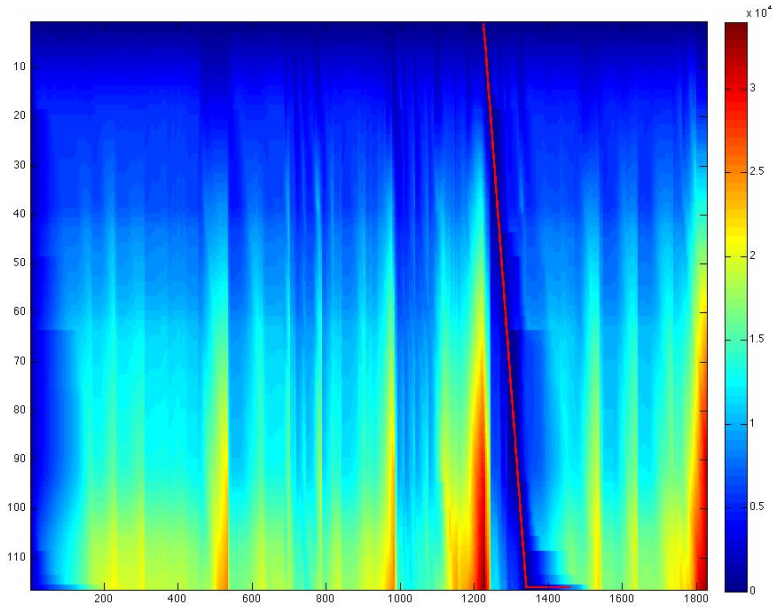


Figura 1. Gràfic de la matriu de distàncies amb recorregut d'inici a fi

2.1.2. Descriptors

Per tal de poder realitzar el Dynamic Time Warping necessitarem extreure uns vectors que contindran les característiques de cada part etiquetada en un fotograma concret, aquests són els descriptors.

Aquests tenen en comú que en tots els casos es calcula el seu valor a partir de la diferència en valor absolut respecte la posició del centroide del seu cap envers la posició del centroide de la extremitat, però en el moment d'obtenir-los tenim en compte altres diferents factors.

Cada matriu de descriptors tindrà una mida de la llargada de fotogrames del vídeo per vint-i-tres, que seran les posicions 'x' i 'y' de totes les parts etiquetades exceptuant el cap. En el cas que hi hagi una part que no aparegui en el fotograma, s'ha decidit posar a zero les seves coordenades.

X1	Y1	X2	Y2	X3	Y3	X4	Y4	X5	Y5	X6	Y6	X7	Y7	X8	Y8	X9	Y9	X 10	Y 10	X 11	Y 11	X 12	Y 12	X13	Y13

Figura 2. Exemple de descriptor

En tot moment agafem com a referència el centroide del cap de l'individu per fer els càlculs però existeix la possibilitat que no aparegui en el fotograma que estem processant, aleshores s'ha decidit agafar l'últim valor vàlid per tal de mantenir la consistència de l'algorisme. Aquest procediment es durà a terme mentre hi hagi una part etiquetada en el fotograma, en cas de no aparèixer cap, es considerarà que en el fotograma no apareix l'individu que estem processant.

Figura 3. Centroide d'una màscara corresponent al cos

Donat que el volum de dades que movem és molt gran i no es poden controlar un per un tots els fotogrames, s'ha inserit una condició en el codi. En cas de tenir una màscara en mal estat agafaríem els valors corresponents al fotograma anterior, ja que la variació d'un fotograma al següent es pràcticament menyspreable, donat que per cada segon del vídeo s'han obtingut 24 fotogrames.

En la següent taula es fa un resum de totes les fórmules aplicades per obtenir els descriptors:

Tipus descriptor	Fórmula utilitzada
1	|distància(centroide(cap),centroide(extremitat))|
2	$\frac{\text{\|distància(centroide(cap),centroide(extremitat))\|}}{\text{àrea bounding box}}$
3	$\frac{\text{\|distància(centroide(cap),centroide(extremitat))\|}}{\text{àrea bounding cap}}$
4	$\frac{\text{\|distància(centroide(cap),centroide(extremitat))\|}}{\Sigma\ (\text{àrees extremitats})}$

Taula 3. Fórmules per el càlcul de cada tipus de descriptor

2.1.2.1. Descriptor tipus 1

En aquest cas, l'operació que emprem simplement és la que tenen en comú els altres 3 mètodes d'obtenció de descriptors, és a dir, la resta entre el centroide del cap i el centroide de l'extremitat en valor absolut. El centroide és el punt que defineix el centre geomètric d'un objecte en un pla i alhora és el punt que està més allunyat de tots els extrems.

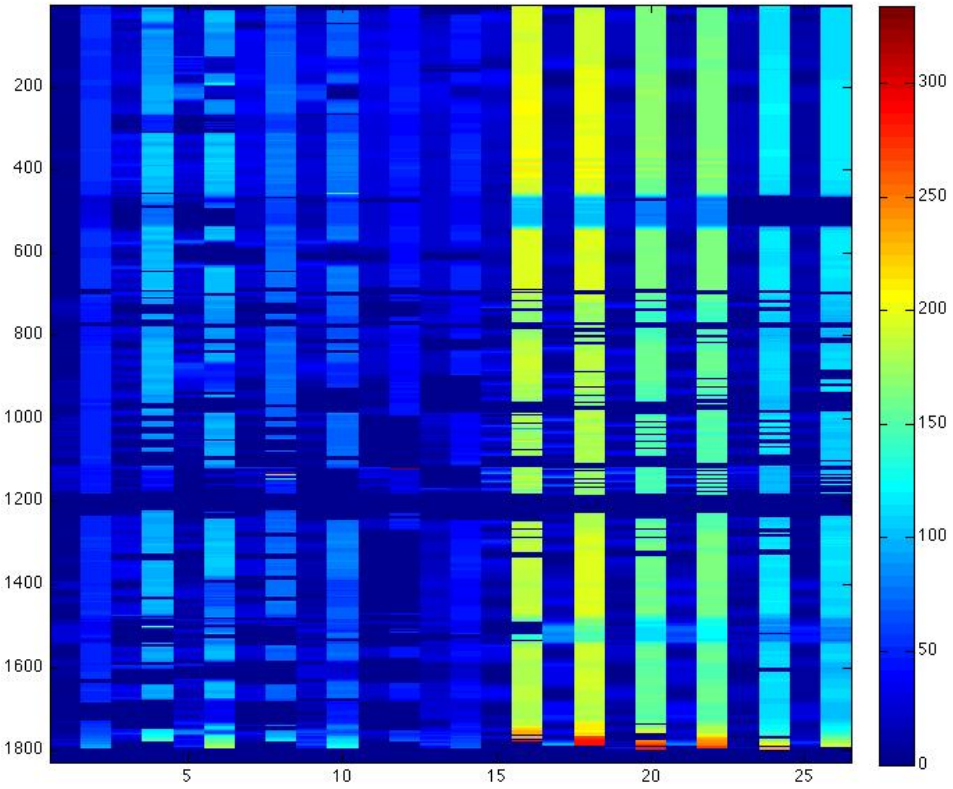


Figura 4. Gràfic dels valors obtinguts del descriptor 1

2.1.2.2. Descriptor tipus 2

En aquest tipus de descriptors emprarem la resta obtinguda del tipus 1 i la dividirem per l'àrea de la bounding box de l'individu. La bounding box, serà l'àrea ocupada en el fotograma per l'individu, tenint en compte les posicions Xmin, Xmax, Ymin, Ymax, és a dir, els extrems de totes les parts etiquetades que veiem en el fotograma. Com que en tot moment l'àrea te una forma rectangular hem aplicat el càlcul de multiplicar la seva base (Xmax-Xmin).

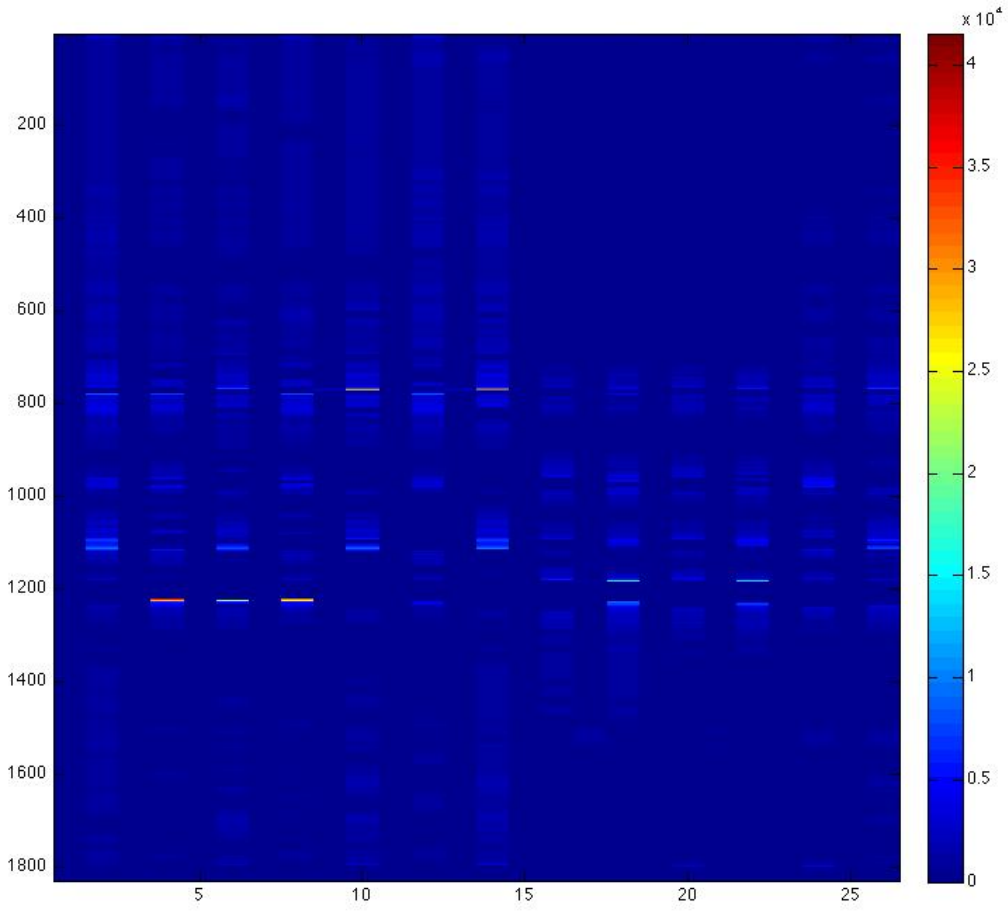


Figura 5. Gràfic dels valors obtinguts del descriptor 2

2.1.2.3. Descriptors tipus 3

Per a realitzar aquest càlcul hem agafat la resta en valor absolut del tipus 1 i s'ha realitzat la divisió del resultat per l'àrea ocupada pel cap. Aquesta àrea ha estat calculada sumant tots els punts de la imatge que contenen la màscara. Un dels punts forts d'aquest mètode és que al normalitzar el descriptor per l'àrea, s'obtenen uns valors més consistents respecte la distància de l'individu i la càmera.

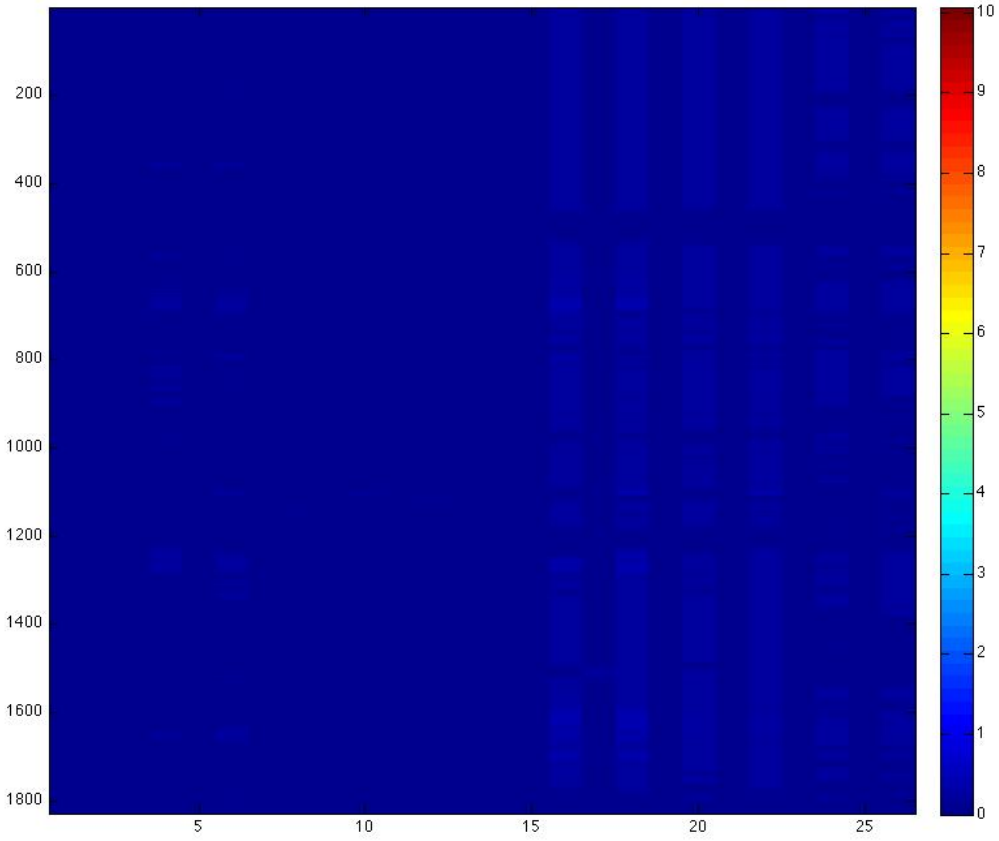


Figura 6. Gràfic dels valors obtinguts del descriptor 3

2.1.2.4. Descriptors tipus 4

En aquest cas, el càlcul que es fa per calcular les distàncies és la divisió de la resta del centroide del cap envers els centroide de la resta d'extremitats dividit pel sumatori de totes les àrees de les extremitats que apareixen. En primer lloc, buscarem les parts que apareixen en un fotograma i procedirem a fer el càlcul de la seva àrea de la mateixa manera que hem calculat l'àrea del cap en els descriptors de tipus 3, després farem la suma de totes elles i farem la seva divisió.

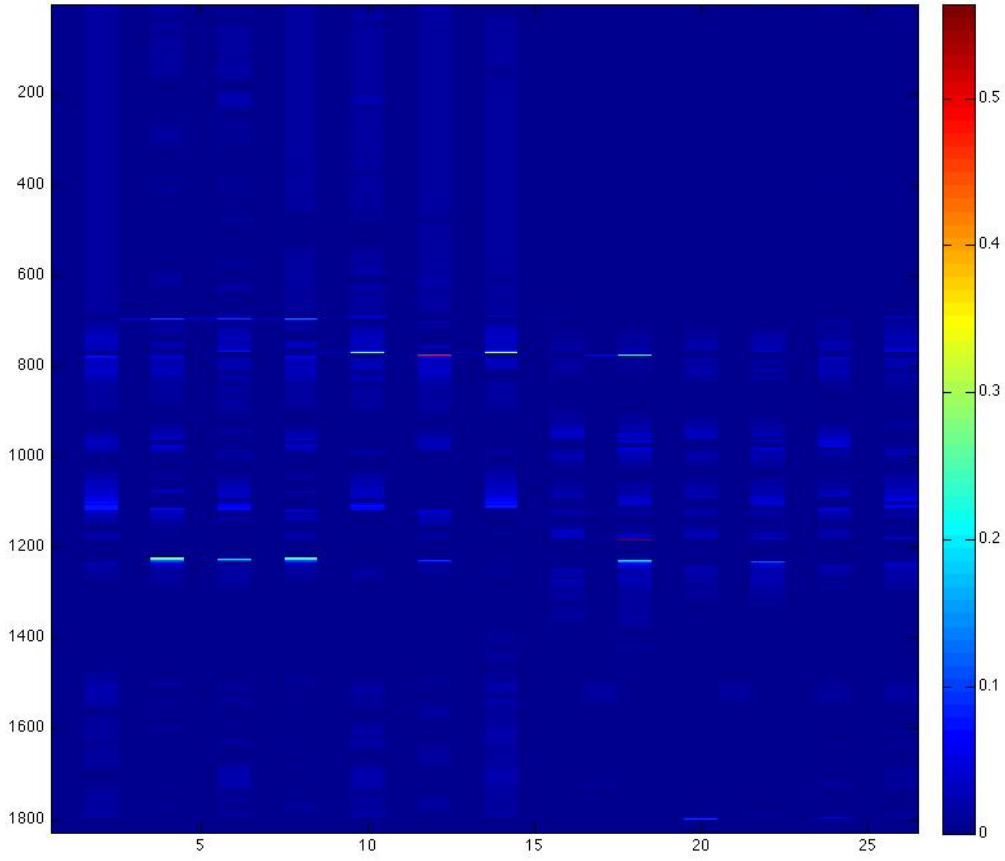


Figura 7. Gràfic dels valors obtinguts del descriptor 4

2.2. Requeriments aplicació

2.2.1. Software

En primer lloc, necessitarem un IDE on programar tots i cadascun dels algorismes que formen part d'aquesta memòria, donat que el llenguatge utilitzat és Matlab emprarem el seu entorn de desenvolupament integrat. En segon, lloc necessitarem dividir els vídeos de mostra en fotogrames (concretament 24 fotogrames per segon), per realitzar aquesta tasca s'ha utilitzat el software per MACOSX anomenat Final Cut.

Per últim, es necessària una aplicació desenvolupada pel meu company de projecte en la fase de la creació de la base de dades escrita en Matlab, aquesta eina és la que ha facilitat la creació de màscares i els seus punts alhora de la recollida de dades de cada extremitat per fotograma. El mèrit del desenvolupament d'aquesta aplicació recau en el meu company de projecte, que va ser capaç d'esbrinar un mètode per poder captar les màscares d'una manera més ràpida, còmode i mecànica.

2.2.2. Hardware

Per tal d'enregistrar els vídeos que nosaltres emprarem per generar la base de dades ha estat necessària una càmera de vídeo i per la realització dels càlculs un ordinador de sobretaula.

2.3. Planificació

Aquest projecte ha tingut una duració de 9 mesos, les diferents activitats dutes a terme estan desglossades en la següent taula:

Activitats	Data inici	Dies	Fi
Creació de la base de dades	01/03/11	121	30/06/11
Presentació de la base de dades	01/07/11	61	31/08/11
Ànalisi problema	31/08/11	30	30/09/11
Disseny i implementació	30/09/11	31	31/10/11
Test DTW	31/10/11	30	30/11/11
Càlcul estadístiques i resultats	30/11/11	46	15/01/12
Escritura memoria	22/12/11	29	20/01/12

Taula 4. Planificació de l'elaboració del projecte

A continuació es mostra un diagrama de Gantt per mostrar la planificació de les diferents etapes del desenvolupament del projecte.

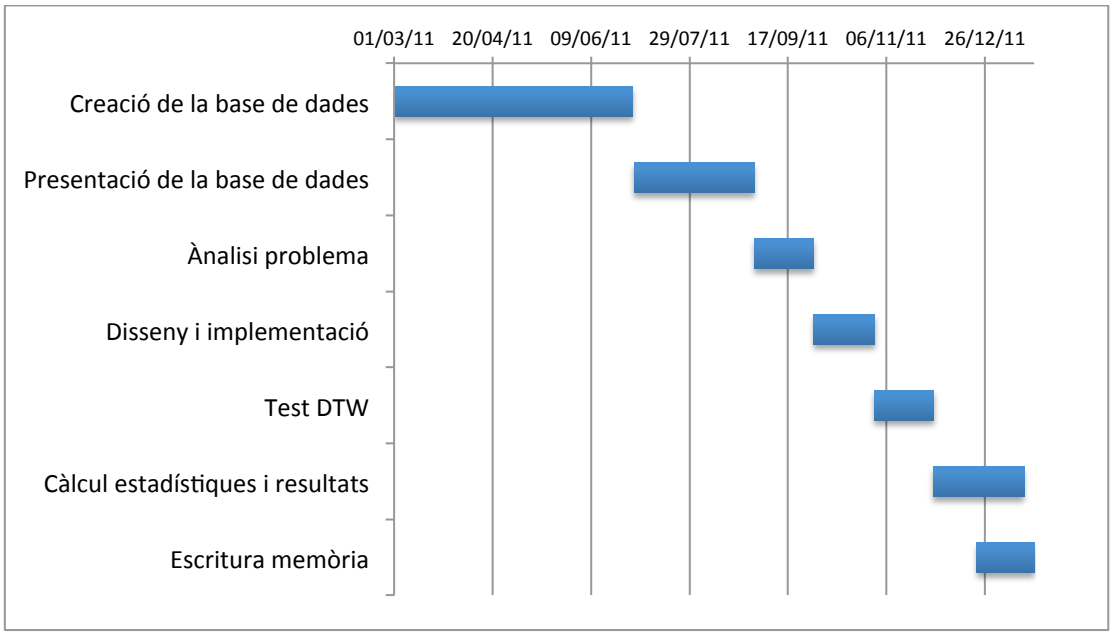


Diagrama 1. Diagrama de Gantt corresponent a la planificació

2.4. Costos

En el moment de calcular els costos s'ha de puntualitzar que en el cas de les hores de feina, s'ha fet un càlcul aproximat donat, que depenen del dia s'hi ha pogut dedicar més o menys hores a l'elaboració del projecte. En la següent taula es mostren les despeses econòmiques desglossades en varis conceptes.

Concepte	Quantitat	Preu	IVA	Preu final
Llicencia individual Matlab	1	600 €	108 €	708 €
Material informàtic	1	1500 €	270 €	1770 €
Hores de treball base de dades (2 usuaris)	320	25 €/h	-	8000 €
Hores de treball	200	25 €/h	-	5000 €
Càmera	1	99 €	18 €	117 €
Total				15195 €

Taula 5. Taula corresponent als costos de la realització del projecte

3. Disseny i implementació

Com s'ha comentat prèviament per la realització del codi s'ha emprat el llenguatge de programació Matlab, degut a que ofereix molts avantatges en el moment de treballar amb matrius, i per tant amb imatges.

3.1. Matlab

MATLAB[9] (abreviatura de MATrix LABoratory), és un software matemàtic que ofereix un entorn de desenvolupament amb un llenguatge de programació propi. Entre les prestacions bàsiques que ofereix s'hi troben: la manipulació de matrius, la representació de dades i funcions, la creació d'interfícies d'usuari (com les que s'han emprat alhora de crear l'eina d'etiquetatge) i la comunicació amb llenguatges.

3.2. Parts del codi

Com en tot el projecte, els comentaris sobre el codi també es troben separats en dos blocs, el primer serà el que comenti els detalls del codi en quan validació de la base de dades i la interpolació. El segon punt es centrarà en la metodologia emprada per obtenir cadascun dels descriptors i l'elaboració de l'algorisme Dynamic Time Warping.

3.2.1. Codi referent base de dades

En aquesta secció s'explicarà el codi realitzat per generar els fotogrames amb els contorns impresos, la interacció entre l'arxiu de dades que conté les dades de les accions i el codi per plasmar en el fotograma l'acció que està realitzant cada individu i per últim el codi per generar la bounding box.

3.2.1.1. Codi de generar els contorns

L'objectiu principal d'aquest algorisme és el de permetre veure a l'usuari si les màscares de la base de dades han sigut correctament etiquetades. El procés que segueix aquest algorisme el podríem resumir de la següent manera: en primer lloc una vegada tenim una màscara, hem de buscar el fotograma al que pertany. Tan bon punt s'ha trobat aquest procedirem a quedar-nos amb el contorn de la màscara, els farem una mica més ample per que es pugui identificar millor, li canviarem el color (per tenir totes les parts identificades en diferents colors) i procedirem a fusionar-lo amb la imatge original. Cal dir que aquesta "fusió" és ni més ni menys que una multiplicació entre dues matrius.

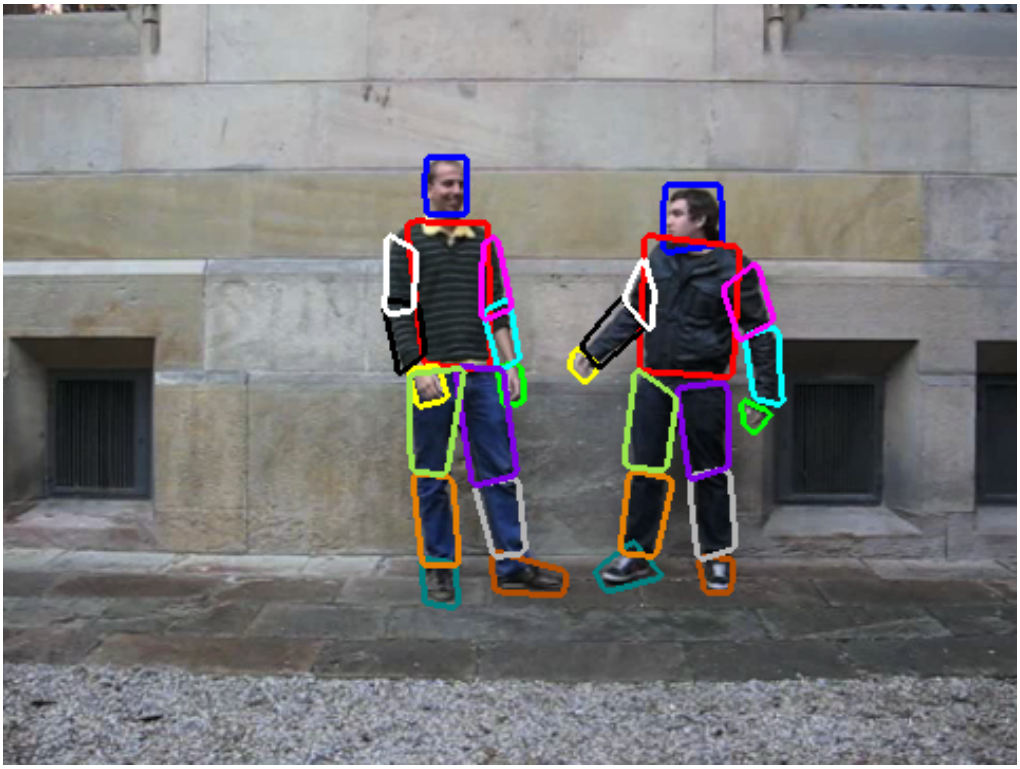


Figura 8. Fotograma on apareixen dos individus amb els contorns de les màscares en color

3.2.1.2. Codi de la implementació de la bounding box

El procediment que segueix l'algorisme requereix l'arxiu que conté els punts que s'han creat en el moment de l'etiquetatge. En aquest cas el que fem és especificar l'usuari del qual estem generant la bounding box i busquem quines són les posicions més situades als extrems de les seves extremitats (explicat en el punt 4.1.6). Una vegada han estat identificades les guardem en una estructura i procedim a fixar-les al fotograma original.

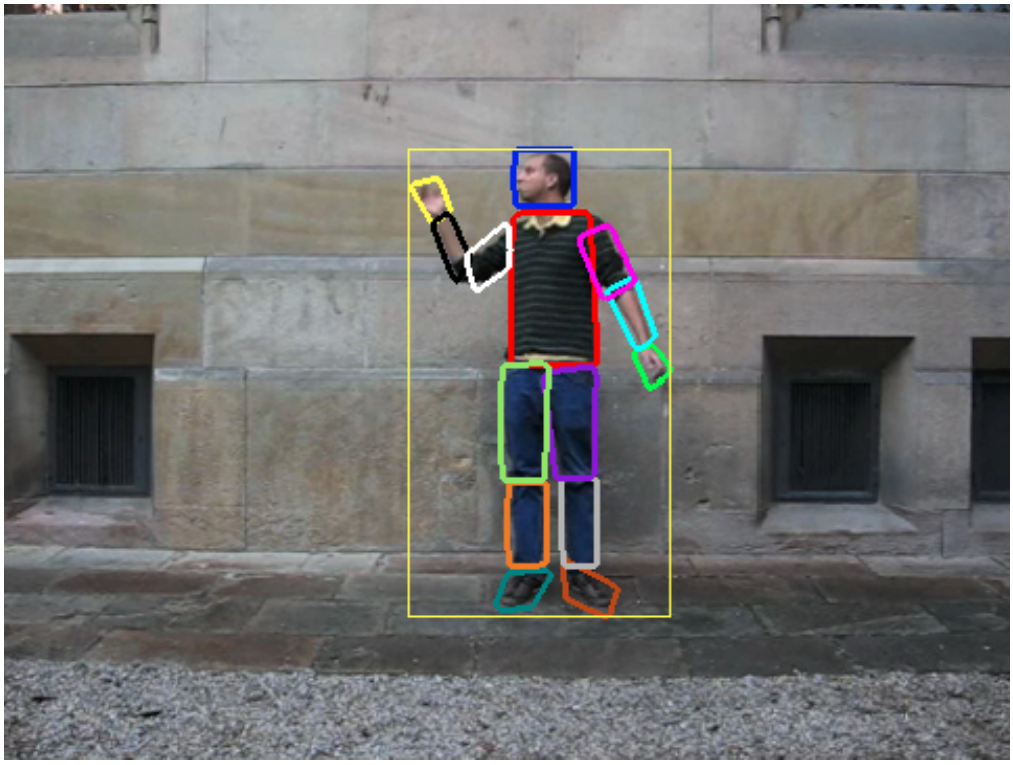


Figura 9. Fotograma on apareix un individu i la seva bounding box

3.2.1.3. Codi de fixar accions en el fotograma

Una vegada hem etiquetat els contorns i la bounding box, procedirem a inserir en la part superior esquerra les accions que estan fent en cada fotograma els individus. Aquí s'ha importat la pàgina de l'arxiu en .xls que conté les dades dels usuaris participants en el vídeo, i per cada fotograma es consulta quina acció està realitzant l'usuari seleccionat, una vegada obtingudes les dades es procedeix a fixar el text. En cas que l'individu hagi realitzat dues accions en el mateix moment també es veurà reflectit en la inserció.

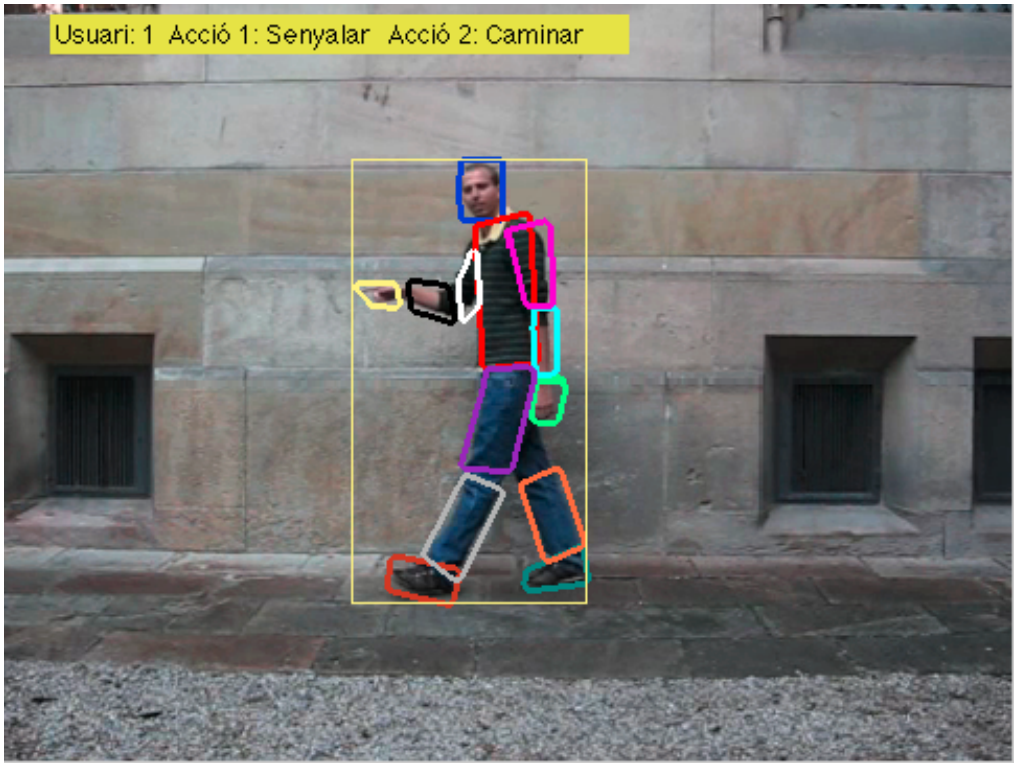


Figura 10. Fotograma amb les dues accions que està realitzant l'individu etiquetades

3.2.2. Codi referent al reconeixement d'accions

En primer lloc parlarem del codi de l'algorisme del Dynamic Time Warping, i seguidament descriurem algunes de les puntualitzacions sobre les operacions realitzades en el moment d'obtenir els descriptors

3.2.2.1. Dynamic Time Warping

Com ja s'ha explicat prèviament aquest algorisme, ens centrarem en el codi. En primer lloc el que es crea és una matriu amb el nombre de línies de la seqüència de la que volem intentar identificar algun patró d'acció i el nombre de columnes serà el nombre de files de la seqüència que emprarem com a mostra. Una vegada fet això es procedeix a calcular la distància euclidiana entre els termes de les dues matrius.

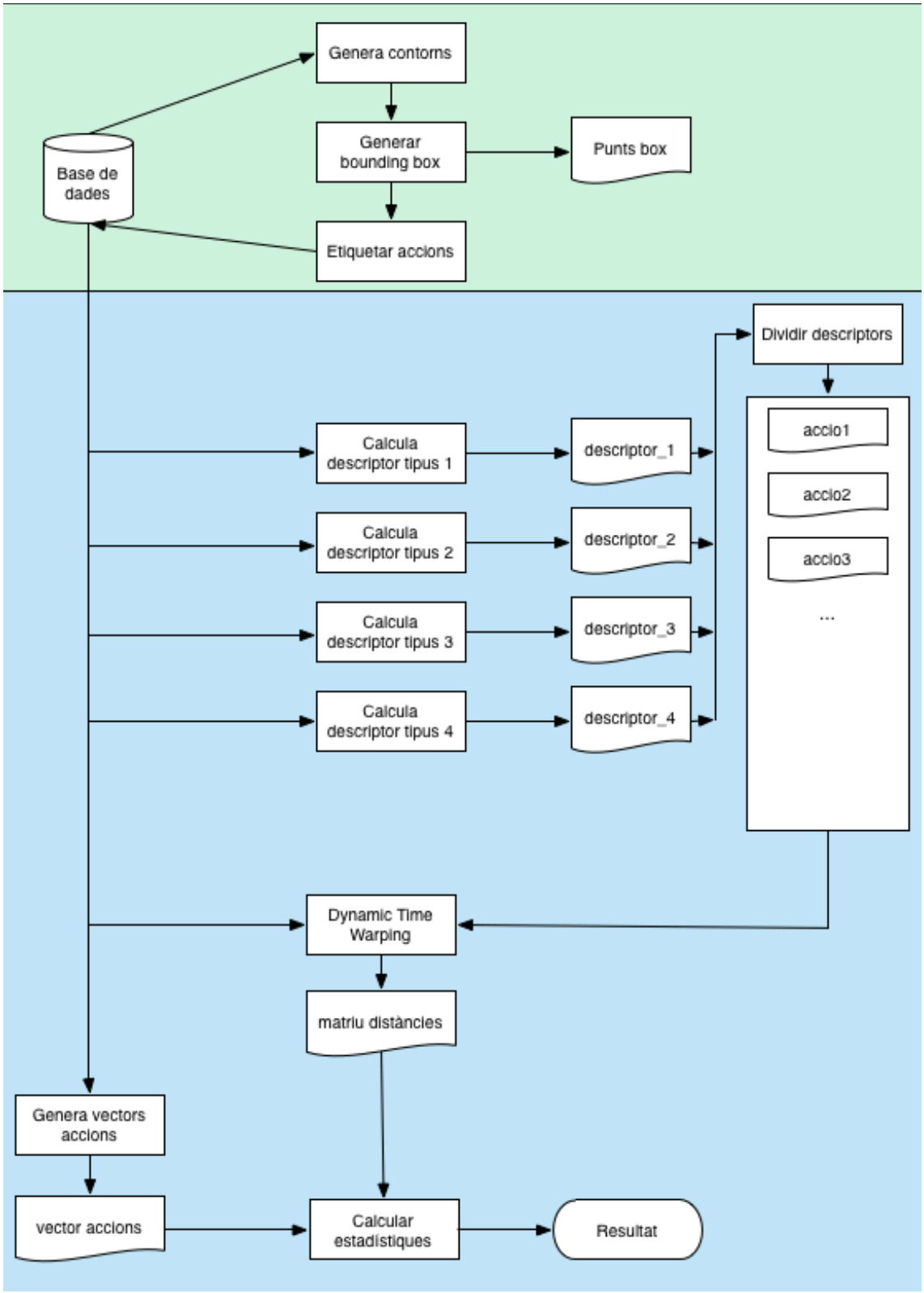
Quan ja ha estat generada la primera matriu, es genera una segona matriu amb les mateixes dimensions que l'anterior on la primera línia estarà plena de zeros, i a cada element de la primera li sumarem el mínim de les posicions superior, esquerra i superior esquerra.

```
function [Dist,D,k,w]=dynamic_time_warping(t,r)

[rows,N]=size(t);
[rows,M]=size(r);

for n=1:N
   for m=1:M
       d(m,n)=pdist2(t(:,n)',r(:,m)','euclidean');
   end
end

D=zeros(size(d));
D(1,1)=d(1,1);

for n=2:M
    for m=2:N
        D(n,m)=d(n,m)+min([D(n-1,m),D(n-1,m-1),D(n,m-1)]);
    end
end
```

Codi en Matlab de Dynamic Time Warping

3.3. Diagrama del projecte

En el següent diagrama de flux es mostra la estructura general del projecte, seguidament explicarem quins són els fitxers que requereix cada algorisme.



3.3.1. Contingut de la Base de dades

En la base de dades hi ha guardats els vídeos originals dividits en fotogrames, que seran als que hi fixarem els contorns de les màscares, la bounding box de cada individu i les accions. També hi tindrem totes les màscares obtingudes a través de l'etiquetatge ben organitzades. Les accions estaran acotades en un arxiu en .xls, on veurem el fotograma d'inici i fi de cada acció duta a terme per cada individu. La estructura que conté els punts de la bounding box i que s'ha generat al fer el seu càlcul també serà guardada a la base de dades.

Aquí també hi seran els vectors d'accions, generats a partir del mateix .xls que hem nombrat abans. D'aquest en tindrem un de cada acció i cada usuari, per cada vídeo de mostra i tindran una mida del número de fotogrames totals que el conformen. La seva composició serà simple, en els fotogrames que l'individu estigui realitzant una acció determinada, al seu vector corresponent hi posarem un 1, en cas de no estar realitzant l'acció hi posarem un 0.

Hi tindrem emmagatzemats també tots els descriptors dels diferents tipus de cada usuari i per cada vídeo, i per últim serà el lloc on estaran guardades totes les matrius de distancies creades amb el Dynamic Time Warping, per tal de poder calcular les estadístiques finals dels nostres experiments.

3.3.2. Algorismes de càlcul de descriptors

Aquests algorismes seran els encarregats de generar els descriptors que posteriorment seran emprats en el Dynamic Time Warping per provar si es pot reconèixer el patró d'una acció en una seqüència desconeguda.

Requeriran les màscares de la base de dades, donat que el seu càlcul es fonamenta en l'extracció del centroide de cada màscara i també en el càlcul de la seva àrea. Una vegada obtinguts els descriptors de les seqüències completes seran subdividits per tal de tenir una mostra exacte de l'acció en el moment que esta esdevenint, i que serà emprada com a patró a reconèixer en el DTW.

3.3.3. Càlcul d'estadístiques

Per el càlcul d'estadístiques necessitarem la matriu de distàncies obtinguda en el Dinàmic Time Warping, i els vectors d'accions (explicats en l'apartat 3.3.1), una vegada obtinguts els resultats seran mostrats en una taula. L'explicació detallada del procediment que es segueix en l'obtenció de estadístiques està en el punt 5.1.

3.4. Problemes sorgits durant l'elaboració del codi

En un principi van aparèixer problemes provocats pel desconeixement total amb el llenguatge que s'ha desenvolupat el projecte, donat que en tot el transcurs de la carrera no s'ha utilitzat. Això va comportar un període d'aprenentatge que va endarrerir de bon principi el desenvolupament del projecte.

En alguns casos es va tractar d'emprar comandes, com per exemple el *parfor*, que facilita el llenguatge per tal d'agilitzar les operacions aprofitant els múltiples nuclis del processador de l'equip i els resultats no van ser del tot satisfactoris, ja que al realitzar-se càlculs en paral·lel es perdien dades.

Per últim alhora d'agafar les màscares i intentar fer la interpolació ens vàrem adonar que en primer lloc el format escollit inicialment no era el correcte (explicat en el punt 4.1.3.), i en segon lloc que seria molt útil guardar els punts de la màscara ja que sinó el procés a seguir per crear els fotogrames interpolats era bastant més difícil perquè es tenien que agafar les dades directament de la imatge.

4. Base de dades

4.1. Creació de la base de dades i extracció de les mostres

En primer lloc m'agradaria comentar que aquesta part no ha estat realitzada íntegrament per mi, seria una falsedat atribuir-me tot el mèrit. Hem estat dues les persones que ens hem encarregat de la realització de la base de dades. Cadascun ens vàrem fer càrrec d'etiquetar manualment la totalitat de vídeos, a més a més, l'eina emprada per realitzar l'etiquetatge de les parts de l'individu no va ser desenvolupada per mi, sinó que va ser cedida per part seva.

Per la creació de la base de dades vàrem agafar els vídeos i van se dividits en imatges per tal d'obtenir-ne 24 mostres per segon. En segon lloc es va decidir que donat que seria una tasca faraònica realitzar l'etiquetatge complet dels vídeos, vam arribar al consens d'etiquetar un de cada dos fotogrames ja que la variació entre un i l'altre és mínima i després realitzar una interpolació mitjançant el fotograma anterior i el posterior per tal d'obtenir el que no ha sigut etiquetat.

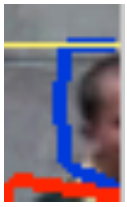
Podem dividir la fase de creació de la Base de dades en 3 parts:

- Etiquetatge de les parts del cos.
- Etiquetatge de les accions.
- Validació de les dades creades.

4.1.1. Protocol d'etiquetatge de les parts del cos

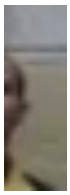
Alhora de decidir si s'ha d'etiquetar una part del cos o no, es van donar com a vàlides les parts de les que veiem un cinquanta per cent o més, ja que sinó podríem agafar mostres massa petites, cosa que provocaria dades errònies en el moment de realitzar els experiments. Al etiquetar les parts també es va tenir en compte deixar un petit marge per tal de tenir una màscara neta.

A continuació es mostra un exemple de cada tipus d'etiquetatge: correcte, en el que podem veure més del cinquanta per cent de la part, i un etiquetatge incorrecte.



En aquest cas, com es pot observar en la imatge no es veu el cinquanta per cent de la part que estem etiquetant i per tant, no s'hauria d'haver etiquetat aquesta part de l'individu, per tant és un clar exemple de mal etiquetatge.

Com es pot observar en la imatge situada a la dreta d'aquestes línies, la totalitat de les imatges que conté la base de dades han estat etiquetades seguint aquest procediment ja que es compleix la regla d'observar més del 50% de la imatge.



Aquí es posa un exemple d'una part que correctament s'ha decidit no etiquetar, donat que només podem veure el quaranta per cent de la seva totalitat, aquestes parts no han estat etiquetades per tal de no interferir negativament en els experiments duts a terme.

4.1.2. Procediment d'etiquetatge de les parts del cos

Una vegada tenim clar el protocol a seguir, en l'aplicació cal especificar el codi de la part que volem etiquetar (en la casella corresponent), també s'haurà d'especificar a quin individu pertany la part que estem etiquetant.

Els individus poden dividir-se en:

- Individus principals: apareixen d'inici a fi en gairebé totes les mostres.
- Individus secundaris: apareixen aproximadament quan ha transcorregut la meitat del vídeo, i participen amb l'individu principal ens les accions que requereixen la interacció entre dos o més usuaris.
- Altres individus: en algun dels vídeos apareixen persones que no formen part de la coreografia estipulada en el vídeo, però també s'han tingut en compte i han estat etiquetades.

Cada individu és etiquetat segons el seu paper en el vídeo, és a dir, a l'individu principal li mantindrem els identificadors de les parts, però als altres individus participants se'ls duplicarà, triplicarà, etc... els seus identificadors per tal de poder diferenciar-los. Així doncs resultaria que si per exemple, estem etiquetant el peu esquerra de l'individu secundari del vídeo l'hauríem d'etiquetar introduint com identificador de part "1010".

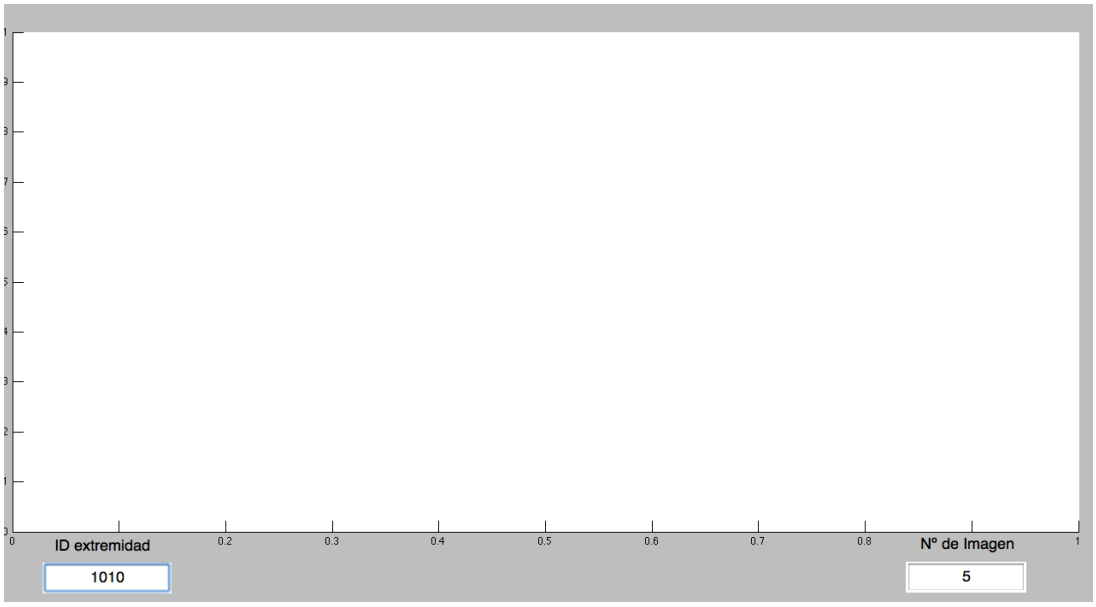


Figura 11. Aspecte de l'aplicació per etiquetar parts en fotogrames

Les diferents parts de l'individu han estat dividides i codificades de la següent manera:

Part	Identificador part
Cap	1
Tronc	2
Mà esquerra	3
Mà dreta	4
Avantbraç dret	5
Avantbraç esquerra	6
Braç dret	7
Braç esquerra	8
Peu dret	9
Peu esquerra	10
Cama dreta	11
Cama esquerra	12
Cuixa dreta	13
Cuixa esquerra	14

Taula 6. Parts etiquetades i el seu corresponent identificador

Mentre és realitzat l'etiquetatge dels individus es va creant una estructura on guardem per cada màscara extreta les següents dades:

- Identificador part
- Usuari
- Número del fotograma
- Els 4 punts que delimiten els vèrtex de la màscara.

Aquestes dades seran emprades en un futur en diferents algorismes, com per exemple alhora de generar les bounding box de cada individu, ja que com em comentat prèviament, tenim relacionats els quatre punts que delimiten la màscara, el fotograma al que pertoquen i l'usuari al que pertanyen.

4.1.3. Format escollit per emmagatzemar les imatges

Per emmagatzemar les imatges extretes el format escollit va ser el ".jpg" però en el moment de realitzar la interpolació ens vam adonar que va ser una decisió errònia donat que el format ".jpg" exerceix una gran compressió a les imatges, cosa que implica una pèrdua de qualitat considerable i fa aparèixer interferències en les imatges. Aleshores es va decidir emprar el format ".bmp" ja que no aplica cap

compressió a la imatge obtinguda, i permet mantenir la qualitat màxima de la imatge fent així més fiables les mostres. L'únic inconvenient que comporta emprar aquest tipus de format és que al no tenir cap tipus de compressió aplicada les dades ocupen cinc vegades més espai en disc dur.

4.1.4. Interpolació

Amb l'estructura obtinguda mentre realitzàvem l'etiquetatge de les parts dels individus, el que farem serà recorre en busca del fotograma que no existeix, i en cas de trobar-lo buscarem la relació entre els punts de cada màscara i mirarem la distància que hi ha entre ells, per quedar-nos amb la mínima.

Una vegada obtinguda donarem com ha posició interpolada per cada punt la resultant de l'operació:

Posició interpolada = (Posició anterior X 0.5) + (Posició posterior X 0.5)

Aquest valor serà guardat en una nova estructura, idèntica en quan a composició a la creada al generar les màscares, però aquest cop contindrà les dades de cada fotograma, tan els interpolats com els introduïts a partir de l'etiquetatge. Alhora serà generada la màscara corresponent als punts calculats.

Existeix la possibilitat que hi hagi un salt de més d'un fotograma, en aquest cas el que farem serà ignorar les dades fins a trobar unes de noves, ja que el que en realitat passa és que la part ha desaparegut visualment del fotograma i no cal aplicar cap tipus d'interpolació.

4.1.5. Etiquetatge de les accions

Les accions que tractarem d'identificar prèviament han estat guardades en un arxiu de dades. Aquestes també han estat acotades manualment fotograma a fotograma.

A cada acció li correspon un identificador:

Acció	Identificador
Saludar	1
Senyalar	2
Aplaudir	3
Senyalar	4
Ajupir-se	5
Saltar	6
Caminar	7
Córrer	8
Donar la mà	9
Abraçar	10
Donar-se dos petons	11
Barallar-se	12
Altres accions	13
Ninguna acció	14

Taula 7. Accions etiquetades i el seu corresponent identificador

Cal assenyalar que en el moment de l'etiquetatge de les accions s'ha tingut en compte com a fotograma d'inici l'instant on s'inicia el moviment d'una part d'un individu que conclou amb la realització de l'acció, és a dir, en el cas de l'acció de senyalar hem considerat que l'acció de senyalar de l'individu s'inicia en el moment que aquest comença a desplaçar el seu braç cap una direcció, i no en el moment que el seu dit índex senyala alguna cosa. S'ha decidit fer-ho d'aquesta manera perquè no tenim accions molts semblants, que puguin conduir cap a una confusió.

En el cas de les accions on participa més d'un individu el que s'ha tingut en compte es quan cadascun dels individus inicia l'acció i no el moment en que els dos individus la estan realitzant, així que és possible veure que un dels dos usuaris ha començat l'acció i l'altre encara no.

4.1.6. Generant la bounding box

Per generar la bounding box necessitarem en primer lloc disposar de l'estructura que s'obté al realitzar la interpolació i que conté tots els punts corresponents als vèrtex de cada màscara.

El procediment a seguir és senzill, en primer lloc intentarem identificar totes les màscares corresponents a un fotograma i que alhora pertanyin a un usuari, una vegada fet això, procedirem a fer una cerca sobre quins són els valors situats més a l'extrem, és a dir, el punt amb la posició més alta (Ymax), el que té la posició més baixa (Ymin), la posició situada més a l'esquerra (Xmin) i la situada més a la dreta (Xmax).

Una vegada tenim els punts que delimiten els extrems en l'espai de l'individu, es procedeix a traçar les rectes entre els punts per delimitar la caixa, també calcularem l'àrea que ocupa.

4.2. Validació de les dades creades

Per validar el correcte etiquetatge de la base de dades al complet s'han generat uns vídeos a partir dels fotogrames obtinguts, que han sigut lleugerament alterats per comprovar que totes les màscares han estat correctament etiquetades, per observar que les accions han estat correctament delimitades en el temps, i per últim amb l'ajut dels contorns generats per validar les màscares veure com la bounding box està correctament situada en els punts. Dos d'aquest vídeos es troben al CD que s'adjunta amb aquest projecte.

4.3. La base de dades en xifres

Per veure el volum de dades creat per a realitzar els experiments, a continuació és mostren algunes xifres significatives:

- Nombre de vídeos: 10

- Fotogrames per vídeo: de 1751 a 1964

- Màscares per vídeo (En primer lloc tenim les màscares etiquetades manualment i entre parèntesi s'observen les que s'han obtingut a través de la interpolació):

    - Vídeo del Actor 1: 13631 (27262)

    - Vídeo del Actor 2: 11954 (23909)

    - Vídeo del Actor 3: 15127 (30254)

    - Vídeo del Actor 4: 12582 (25166)

    - Vídeo del Actor 5: 13015 (26030)

    - Vídeo del Actor 6: 13827 (27654)

    - Vídeo del Actor 7: 18926 (37852)

    - Vídeo del Actor 8: 12300 (22160)

    - Vídeo del Actor 9: 21970 (43940)

    - Vídeo del Actor 10: 11080 (22160)

- Nombre màxim de parts de l'individu etiquetades en un fotograma: 14

- Nombre mínim de parts de l'individu etiquetades en un fotograma: 1

- Nombre mínim d'usuaris que apareixen en un vídeo: 2

- Nombre màxim d'usuaris que apareixen en un vídeo: 5

- Màscares etiquetades manualment: 144410

- Volum total de la base de dades: 288827

5. Resultats

Per a la realització dels experiments amb el Dynamic Time Warping i les seqüències d'accions, s'ha decidit no emprar en el reconeixement accions del mateix usuari, és a dir, quan intentem identificar una acció mitjançant l'algorisme no ho farem amb les mostres obtingudes del mateix individu, sinó que emprarem les accions fetes per altres usuaris en altres vídeos. És un mètode de provar l'efectivitat de l'algorisme de reconeixement ja que el patró que enfrontarem amb la seqüència completa és totalment desconegut per aquesta. Així doncs si obtenim resultats positius podrem validar la consistència dels algorismes emprats.

5.1. Protocol de validació del reconeixement

Per validar que la acció que volem reconèixer apareix en la seqüència completa on la volem identificar el procediment a seguir és el següent:

- En primer lloc hem de tenir el vector d'uns i zeros per tal de tenir una referència exacta de quan apareix l'acció que estem buscant en el vídeo. Aquest vector ha estat generat a través de l'excel on apareix acotada cada acció i per tant podem considerar-lo com el cas ideal.

- En segon lloc necessitarem obtenir un vector, també d'uns i zeros, generat a partir de la última línia obtinguda de emprar el Dynamic Time Warping entre el descriptor de l'acció que volem identificar i la seqüència completa (apartat 5.1.2).

- Tan bon punt disposem dels dos vectors, l'operació que es realitza per calcular el tant per cent d'èxit en el reconeixement és:

$$\text{Vector intersecció} = \text{vector original acció} \times \text{vector obtingut}$$
$$\text{Vector unió} = \text{vector original acció} + \text{vector obtingut}$$

$$\text{Percentatge d'encert} = \left(\sum(\text{vector intersecció})*100\right) / \sum(\text{vector unió})$$

5.1.1. Obtenció del valor de tall òptim

Per a l'obtenció del valor de tall òptim s'han realitzat tests amb tots els valors del rang que engloben tots els resultats obtinguts en la matriu de distàncies, es a dir, s'ha provat un per un cada valor i s'ha comparat amb els anteriors, en cas de ser el millor obtingut fins el moment hem decidit guardar-lo i agafar-lo com a nova referència. Per cada valor s'ha creat un vector amb les possibles localitzacions de l'acció en la seqüència original.

5.1.2. Creació del vector a partir de la matriu de distàncies

Per cada valor de tall el procediment per emplenar el vector obtingut consistirà en recorre la última línia de la matriu de distàncies. Una vegada trobem un valor que és més petit (A en la imatge) o igual al valor de tall òptim el que fem es trobar el fotograma d'inici i així acotar el 'camí' de l'acció en la matriu. Una vegada obtenim el fotograma d'inici (B en la imatge), on segons hem detectat s'origina l'acció, procedim a emplenar el vector obtingut de l'acció, inserint uns en les posicions que van des del fotograma detectat fins on s'origina (inici).

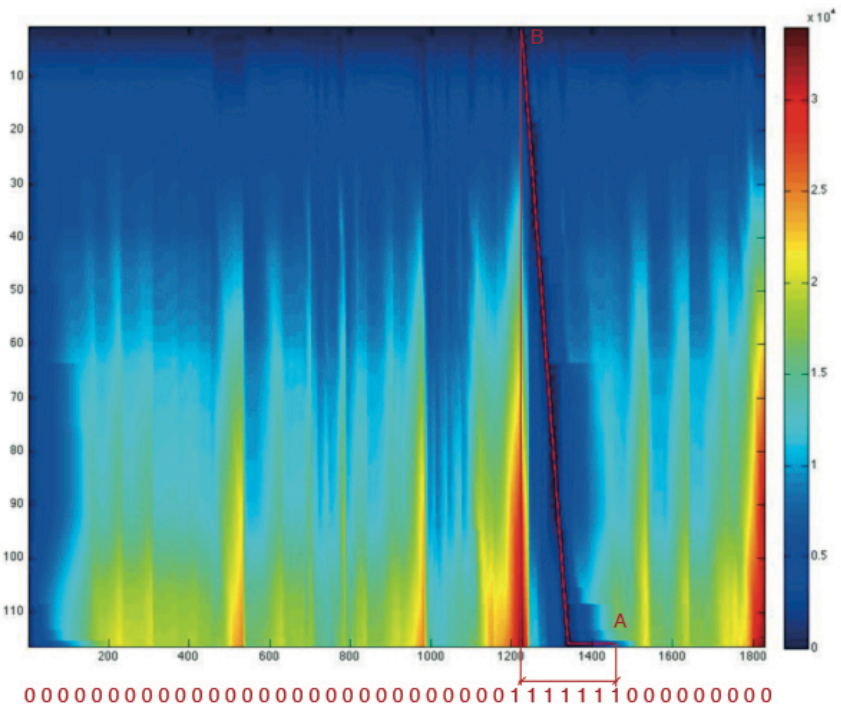


Figura 12. Gràfic d'una matriu de distàncies amb el camí identificat

Inicialment el que es va pensar va ser el d'unir tots els vectors obtinguts, però això hagués comportat una propagació d'errors que ens hagués acabat perjudicant els percentatges d'èxit, però es va observar que va ser un mal plantejament. Així doncs, es va decidir que, en cas de tenir varies mostres de l'acció el que fem és fer el càlcul per cada mostra i obtenir-ne una mitjana de tots els percentatges obtinguts, aquest mètode és emprat perquè en cas de tenir una mostra de l'acció on obtenim un percentatge de reconeixement baix, no perjudiqui al resultat final ja que amb les altres mostres es poden haver obtingut percentatges més elevats.

Exemple:

Vector original :

0 0 0 0 0 0 0 0 0 0 0 1 1 1 0 0 0 0 1 1 1 0 1 0

Vector obtingut 1:

0 0 0 1 0 0 0 0 0 1 1 1 1 1 0 0 0 1 1 1 0 0 1 0

Vector obtingut 2:

1 1 1 0 0 0 0 1 1 1 0 0 0 1 1 1 0 0 0 1 1 0 0 0

Vector obtingut de la unió dels 2 obtinguts:

1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 0 1 0

Vector obtingut de la intersecció dels 2 obtinguts:

0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 0 0 1 1 0 0 0 0

Si fem la comparació amb tots els vectors alhora tenim que:

(Intersecciox100)/Unió = (4x100)/18 = 22%

Si la fem individualment observem que:

Vector original amb vector obtingut 1 : (6x100)/10 = 60%

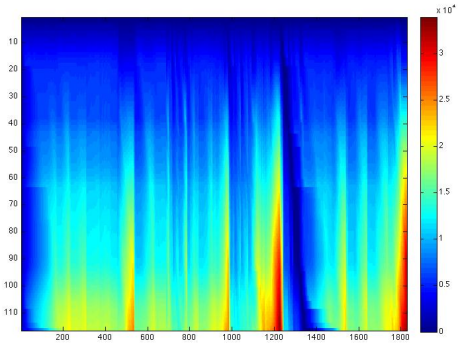
Vector original amb vector obtingut 2 : (4x100)/18 = 22%

Mitjana entre els dos tant per cents: (60+22)/2 = 44%

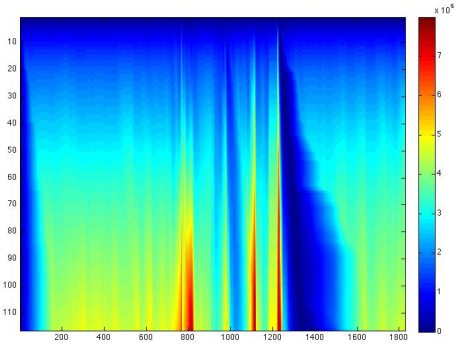
5.1.3. Proves realitzades

Per a poder assegurar la consistència del mètodes emprats per a l'obtenció dels resultats a continuació es mostren diferents representacions gràfiques corresponents a diverses proves prèvies per validar l'efectivitat dels algoritmes.

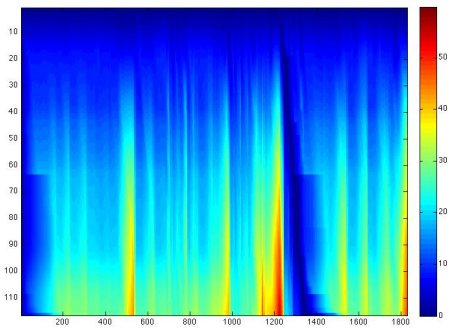
En primer lloc, les quatre imatges següents mostren la mateixa acció, detectada sobre el mateix vídeo, per tant s'hauria de poder identificar clarament l'acció en el gràfic.



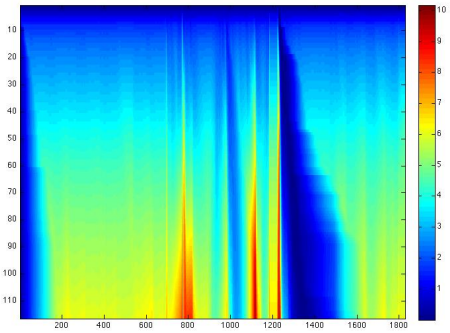
Tipus 1



Tipus 2



Tipus 3

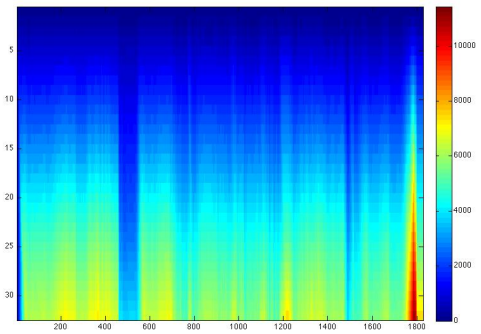


Tipus 4

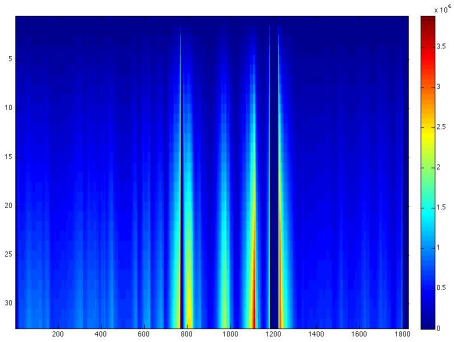
Es pot comprovar clarament l'efectivitat dels algorismes donat que l'acció que es tracta de detectar es realitza entre els fotogrames 1226 i 1341, coincidint amb la part més blava dels gràfics (que ens indica l'interval on és realitzada l'acció).

Una vegada s'ha validat que en la seqüència ens detecta l'acció duta a terme pel mateix individu, cal validar el procediment amb la mateixa acció realitzada per un individu diferent.

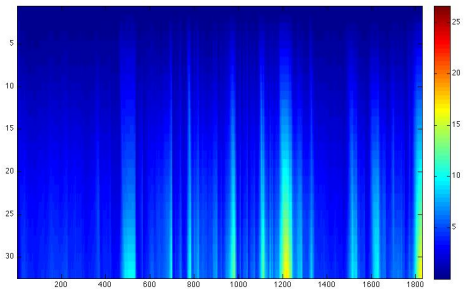
A continuació mostrarem les gràfiques de les matrius de distàncies obtingudes després de provar de detectar l'acció d'ajupir-se realitzada per l'actor 2 en la seqüencia d'accions dutes a terme per l'actor 1.



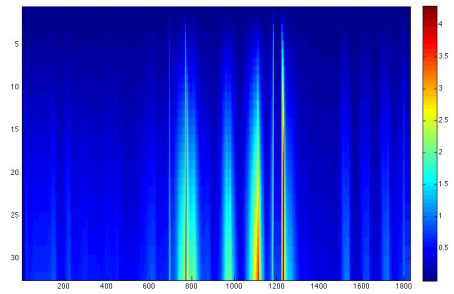
Tipus 1



Tipus 2



Tipus 3



Tipus 4

Es pot observar que a ull nu, l'acció només és identificable en el tipus 1 i 2 (concretament entre els fotogrames 263 i 311), per això necessitem l'algorisme que ens calcula el valor de tall òptim, ja que sinó és possible que no poguéssim donar uns resultats tan precisos com els que es mostraran en l'apartat 5.3.

5.2. Dades de la BD agafades per la validació

Per problemes de disponibilitat de la resta de la base de dades i de temps, es va decidir utilitzar per la realització dels experiments els fotogrames obtinguts a partir dels vídeos de l'actor 1 i de l'actor 2. En total tenim 6 individus diferents realitzant algunes o totes les accions, a continuació es fa una breu descripció de les accions dutes a terme per cada individu.

Vídeo de l'actor 1:

- Individu principal: totes les accions.
- Individu secundari: abraçar, barallar, caminar, donar la mà, donar dos petons.
- Altre individu 1: caminar.
- Altre individu 2: caminar.

Vídeo de l'actor 2:

- Individu principal: totes les accions.
- Individu secundari: abraçar, barallar, caminar, donar la mà, donar dos petons.

Com es pot observar hi ha accions que només es testejaran d'un usuari a un altre, i per tant, si la mostra a comparar és bona tindrem un gran tant per cent d'encert, però pel contrari, en cas de ser una mostra dolenta no tindrem cap alternativa per obtenir un tant per cent millor.

5.3. Taules de resultats

Les caselles que es mostren en negre, són caselles buides ja que l'usuari no fa l'acció en la seqüència original. Les caselles que veiem ressaltades en taronja contenen el percentatge més elevat d'encert d'entre tots els tipus de descriptors.

5.3.1. Taula de reconeixement emprant els descriptors de tipus 1

Individu\Acció	Saludar	Senyalar	Aplaudir	Saltar	Ajupir	Caminar
Vídeo 1 actor 1	9.0909	14.7914	16.4728	3.3273	64.8855	38.1473
Vídeo 1 actor 2						40.2014
Vídeo 1 actor 3						63.1234
Vídeo 1 actor 4						53.1840
Vídeo 2 actor 1	22.9730	39.8844	14.4958	13.7143	17.9775	29.3889
Vídeo 2 actor 2						24.7517

Individu\Acció	Córrer	Abraçar	Donar mà	Donar petons	Barallar
Vídeo 1 actor 1	15.9788	44.2722	30.3222	16.7621	23.4571
Vídeo 1 actor 2		26.3002	17.7384	24.4689	21.4356
Vídeo 1 actor 3					
Vídeo 1 actor 4					
Vídeo 2 actor 1	26.5246	29.1266	10.1455	8.3278	14.3187
Vídeo 2 actor 2		21.3463	22.4168	16.3341	12.1431

5.3.2. Taula de reconeixement emprant els descriptors de tipus 2

Individu\Acció	Saludar	Senyalar	Aplaudir	Saltar	Ajupir	Caminar
Vídeo 1 actor 1	10.4265	3.2491	3.2343	3.6738	12.5690	20.1753
Vídeo 1 actor 2						17.6281
Vídeo 1 actor 3						15.4526
Vídeo 1 actor 4						18.2214
Vídeo 2 actor 1	12.3064	2.7675	4.1250	9.2231	9.7780	12.8733
Vídeo 2 actor 2						16.1292

Individu\Acció	Córrer	Abraçar	Donar mà	Donar petons	Barallar
Vídeo 1 actor 1	2.9174	14.2602	9.1457	8.2403	8.1457
Vídeo 1 actor 2		8.4588	13.5139	9.5741	3.4571
Vídeo 1 actor 3					
Vídeo 1 actor 4					
Vídeo 2 actor 1	13.6645	19.1376	8.2346	8.4302	8.0230
Vídeo 2 actor 2		12.5890	18.2300	7.1470	15.6512

5.3.3. Taula de reconeixement emprant els descriptors de tipus 3

Individu\Acció	Saludar	Senyalar	Aplaudir	Saltar	Ajupir	Caminar
Vídeo 1 actor 1	54.4304	19.2488	43.1373	11	6.6779	49.1629
Vídeo 1 actor 2						60.1234
Vídeo 1 actor 3						40.1299
Vídeo 1 actor 4						27.1538
Vídeo 2 actor 1	50.6024	20.4348	30.4348	15.7377	9.7654	33.1457
Vídeo 2 actor 2						26.7685

Individu\Acció	Córrer	Abraçar	Donar mà	Donar petons	Barallar
Vídeo 1 actor 1	20.1839	30.1376	48.4677	36.7332	30.4218
Vídeo 1 actor 2		12.3342	22.5625	26.3451	12.6398
Vídeo 1 actor 3					
Vídeo 1 actor 4					
Vídeo 2 actor 1	17.9141	9.1014	53.3112	24.7913	13.7456
Vídeo 2 actor 2		16.7991	40.6731	8.3426	30.1225

5.3.4. Taula de reconeixement emprant els descriptors de tipus 4

Individu\Acció	Saludar	Senyalar	Aplaudir	Saltar	Ajupir	Caminar
Vídeo 1 actor 1	17.9688	23.2955	5.7579	6.4252	17.7734	67.4582
Vídeo 1 actor 2						45.2374
Vídeo 1 actor 3						42.1346
Vídeo 1 actor 4						50.6788
Vídeo 2 actor 1	16.4062	40.7407	12.3693	39.7306	6.1657	38.4137
Vídeo 2 actor 2						23.1435

Individu\Acció	Córrer	Abraçar	Donar mà	Donar petons	Barallar
Vídeo 1 actor 1	21.1644	57.2251	60.3001	24.7126	18.4500
Vídeo 1 actor 2		49.3780	12.2892	32.4753	20.1412
Vídeo 1 actor 3					
Vídeo 1 actor 4					
Vídeo 2 actor 1	16.5517	19.0912	22	19.4208	14.7183
Vídeo 2 actor 2		55.1347	30.4584	25.3870	11.4715

5.4. Avaluació dels resultats obtinguts

Com es pot observar a través de les taules, veiem que hi ha un tipus que conté més valors alts respecte dels altres en la detecció de les accions, per tant podem arribar a la conclusió que dels quatre mètodes plantejats el tipus 3 és el millor, seguit molt a prop del tipus 4.

La diferència en el reconeixement entre els tipus 3, 4 i els tipus 1, 2 pot ser deguda a la normalització dels valors dels primer per les diferents àrees de cada part del cos, per tant podríem dir que són factors de pes en els mètodes de reconeixement utilitzats.

Com era d'esperar, el moviment amb millor taxa de reconeixement en tots els casos és el de caminar, donat que tots els individus apareguts en els vídeos el realitzen de forma molt similar tant en velocitat com en moviment de les parts del cos, que és el que analitzem nosaltres. Cal mencionar que si es comparen els individus 3 i 4 del vídeo de l'actor 1 entre ells els percentatges de reconeixement són molt elevats donat que les seves mostres són molt semblants.

S'ha de comentar que en l'acció de barallar, ens esperàvem obtenir resultats més baixos donat que en la realització de l'acció hi ha molts factors que entren en marxa, i cada usuari mou les parts del cos de forma diferent, és a dir, és una acció que no realitzen de forma homogènia tots els usuaris.

Així doncs, un factor determinant en quan al reconeixement d'accions és el moviment de les parts del cos, ja que en les taules, els valors de reconeixement més alts corresponen a les accions que realitzen de manera semblant els diferents individus per a tots els tipus d'obtenció de descriptors que fem servir.

Com a curiositat, s'ha de comentar que les accions de senyalar i donar la mà si són comparades entre si ens donen percentatges d'encert bastant sorprenents, però si és té en compte la ultima conclusió que hem fet en el paràgraf anterior, es conclou que es degut a que les parts del cos és mouen de forma semblant i no es té en compte l'acció de l'altre usuari.

6. Conclusions

En el present projecte s'ha estudiat, desenvolupat, implementat i documentat un mètode de reconeixement d'accions humanes en múltiples vídeos. El nostre algorisme és capaç de detectar i reconèixer les diferents accions que apareixen en les seqüències de vídeo, servint-se només d'una simple mida estadística, amb la que podem mesurar la semblança entre dos gestos.

Com és pot comprovar s'han complert els dos objectius principals del projecte. El primer, realitzar una base de dades lo suficientment gran per poder comparar les mostres i contribuir en aquest camp d'investigació; i el segon, provar i validar quin és el millor mètode d'obtenció de descriptors dels quatre plantejats per al reconeixement d'accions.

Cal citar que en l'aplicació de l'algorisme DTW, cal que existeixi un tram molt semblant al patró que estem buscant, i que amb el mètode d'obtenció del valor de tall de manera estadística s'han trobat resultats bastant més efectius que acotant el valor de tall manualment.

Mentre s'ha estat desenvolupant aquest projecte, hem pogut comprovar que la tasca del reconeixement d'accions en un vídeo és bastant difícil. Actualment és un aspecte pendent en el món del processament d'imatge i vídeo. Si es vol prosseguir en aquest camp, caldrien moltes més dades i ampliar el nombre d'accions registrades, ja que la nostre cerca està limitada només a 11 accions, i encara que tot i haver creat una gran base de dades, no és suficient per assegurar percentatges de reconeixement elevats.

S'aspira a que amb aquest projecte s'hagi pogut contribuir a ampliar els coneixements amb els sistemes de reconeixement i ajudar a aquest camp amb la contribució aportada per la nostre base de dades de mostres.

Com a possibles millores, es podrien desenvolupar projectes on es tingui en compte la distància a la càmera, o la interacció amb altres individus, és a dir, en el cas de les accions on necessàriament han de participar dos individus.

7. Referències

[1] http://en.wikipedia.org/wiki/Dynamic_time_warping

[2] http://www.mendeley.com/research/using-dynamic-time-warping-find-patterns-time-series/

[3] http://www.aaai.org/Papers/Workshops/1994/WS-94-03/WS94-03-031.pdf

[4] http://es.wikipedia.org/wiki/Modelo_oculto_de_M%C3%A1rkov

[5] http://jedlik.phy.bme.hu/~gerjanos/HMM/node2.html

[6] (REF. [Rabiner,83] L.R.Rabiner, S.E.Levinson, M.H.Sondhi: *"On the Application of Vector Quantization and Hidden Markov Models to Speaker-Independent, Isolated Word Recognition"*. The Bell System Technical Journal. Vol.62, Nº.4, April 1983.)

[7] http://www.xbox.com/es-ES/kinect

[8] http://ca.wikipedia.org/wiki/Dist%C3%A0ncia_euclidiana

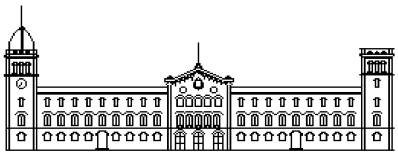
[9] http://www.mathworks.es/products/matlab/index.html

8. Annexos

8.1. Annex 1: Contingut del CD

Dins del CD adjuntat amb la memòria trobarem el següent contingut:

- \vídeo\: Directori amb el vídeo del actor 1 amb els contorns de les mascares, la bounding box i les accions marcades.
- \memoria.pdf : memòria del projecte.
- \dades\: directori amb les màscares corresponents al vídeo del actor 1 i l'actor 2, algunes matrius de distàncies obtingudes al realitzar el DTW sobre les mostres, les estructures generades al etiquetar les màscares, els vectors d'accions, i l'excel amb les accions definides.
- \src\: directori amb el codi dels algorismes desenvolupats per la realització del projecte.



ENGINYERIA TÈCNICA EN INFORMÀTICA DE SISTEMES

UNIVERSITAT DE BARCELONA

Treball fi de carrera presentat el dia de de 2012

a la Facultat de Matemàtiques de la Universitat de Barcelona,

amb el següent tribunal:

Dr. President

Dr. Vocal 1

Dr. Secretari

Amb la qualificació de :