
HUMAN POSE ANALYSIS AND GESTURE RECOGNITION FROM DEPTH MAPS: METHODS AND APPLICATIONS

MIGUEL REYES ESTANY

DEPARTAMENT DE MATEMÀTICA APLICADA I ANÀLISI

UNIVERSITAT DE BARCELONA

27 DE FEBRER DE 2017

SUPERVISOR: DR. SERGIO ESCALERA



OUTLINE

1. Introducció i motivació
2. Reconeixement de la postura basat en descriptors RGB-D
3. Reconeixement de l'activitat basat en mapes de profunditat
4. Descriptor *Spherical Blurred Shape Model*
5. Conclusions
6. Contribucions



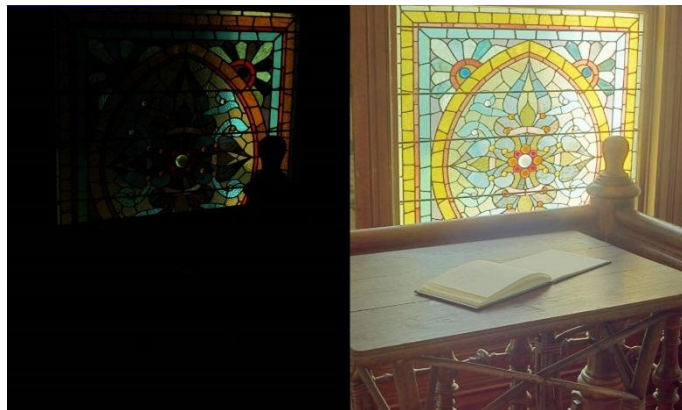
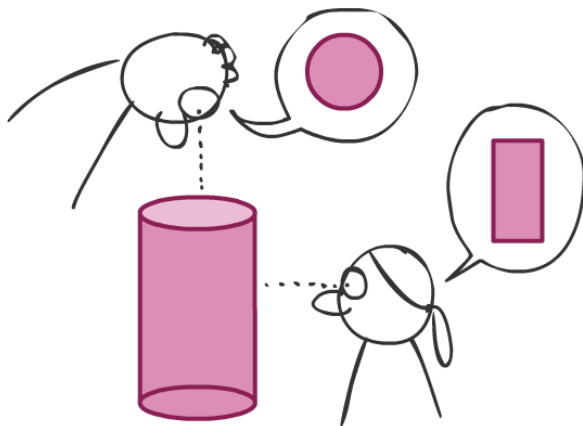
OUTLINE

1. Introducció i motivació
2. Reconeixement de la postura basat en descriptors RGB-D
3. Reconeixement de l'activitat basat en mapes de profunditat
4. Descriptor *Spherical Blurred Shape Model*
5. Conclusions
6. Contribucions



1. TRACTAMENT D'UN PROBLEMA BASAT EN VISIÓ PER COMPUTADOR

- A l'hora d'aplicar conceptes teòrics de la visió per computador trobem interferències i problemes relacionats amb el món que ens envolta.
- Aquest fet dóna lloc a distorsions que contaminen la imatge que s'està analitzant i, en ocasions, pot donar pas a malinterpretacions de l'escena.



1.1 EVOLUCIÓ DE L'ANÀlisi D'UN PROBLEMA DE VISIÓ PER COMPUTADOR BASAT EN IMATGE 2D

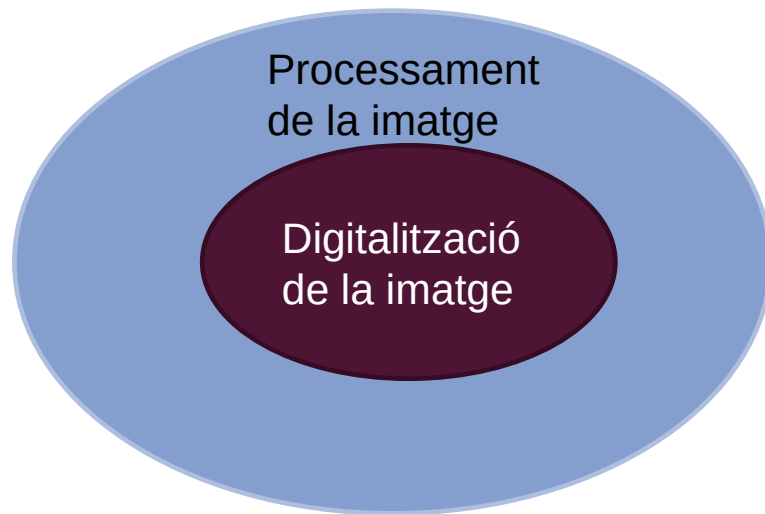
Digitalització
de la imatge

Problemes de percepció
de la imatge

- Es poden representar gràficament les ones electromagnètiques, a diferents longituds d'ona, per tal de ser analitzades.
- Obtenció i discriminització de diferents graus de lluminositat.
- Obtenció de formes, textures i color.



1.1 EVOLUCIÓ DE L'ANÀLISI D'UN PROBLEMA DE VISIÓ PER COMPUTADOR BASAT EN IMATGE 2D



Problemes de percepció de la imatge

- Supressió de soroll tècnic.
- Realçar o minvar detalls de la imatge a través de l'aplicació de filtres no lineals.

Dicot Leaf Epidermis Before and After Image Processing

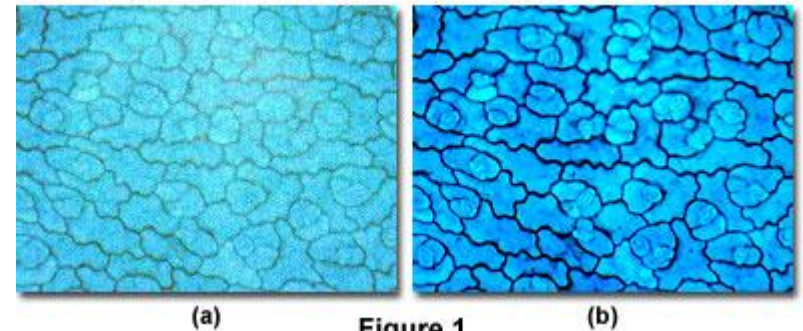
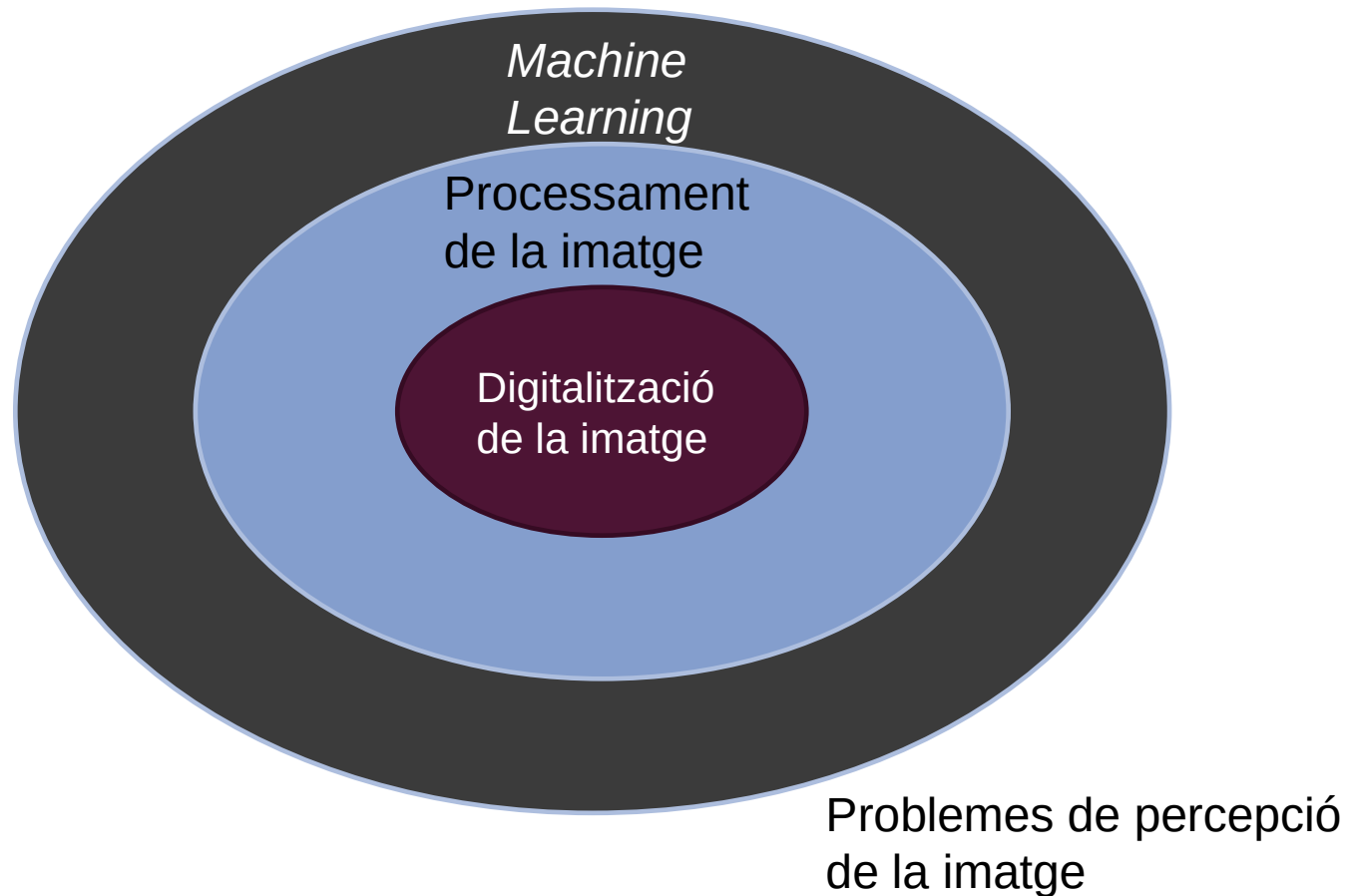


Figure 1



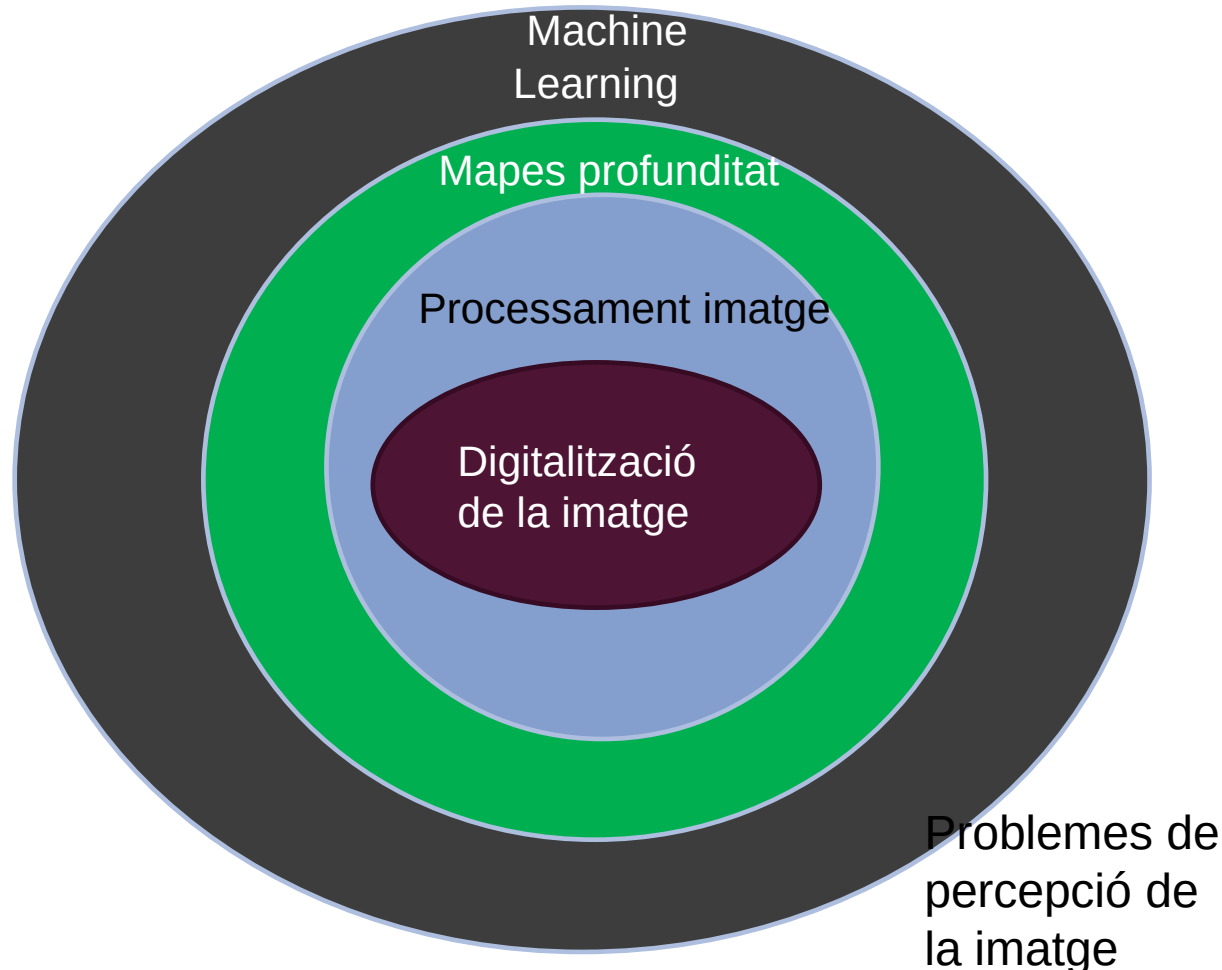
1.1 EVOLUCIÓ DE L'ANÀLISI D'UN PROBLEMA DE VISIÓ PER COMPUTADOR BASAT EN IMATGE 2D



- Generació de models geomètrics, probabilístics i lògics per a la detecció de patrons
- Classificació i segmentació de la imatge

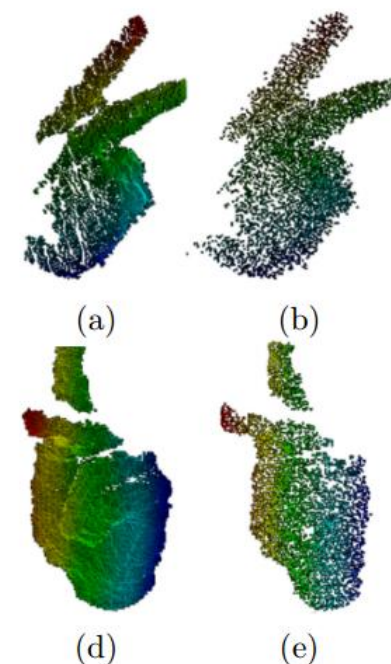
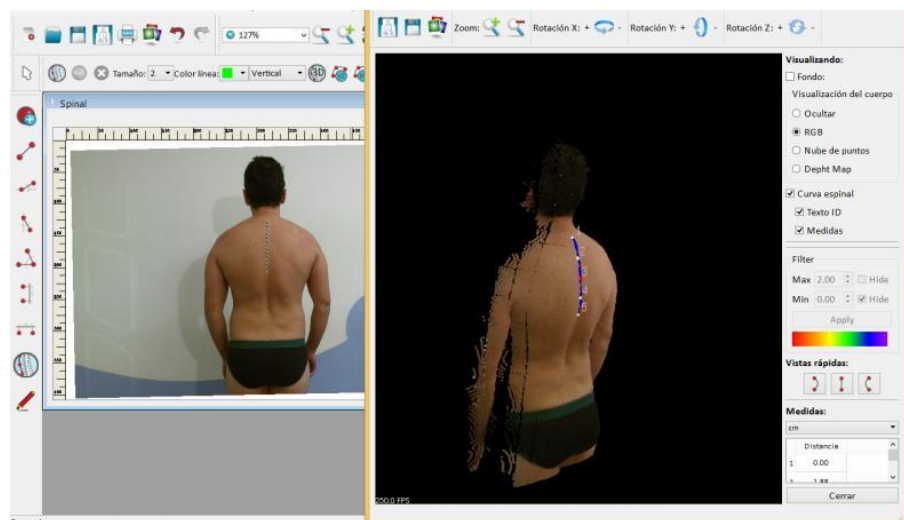
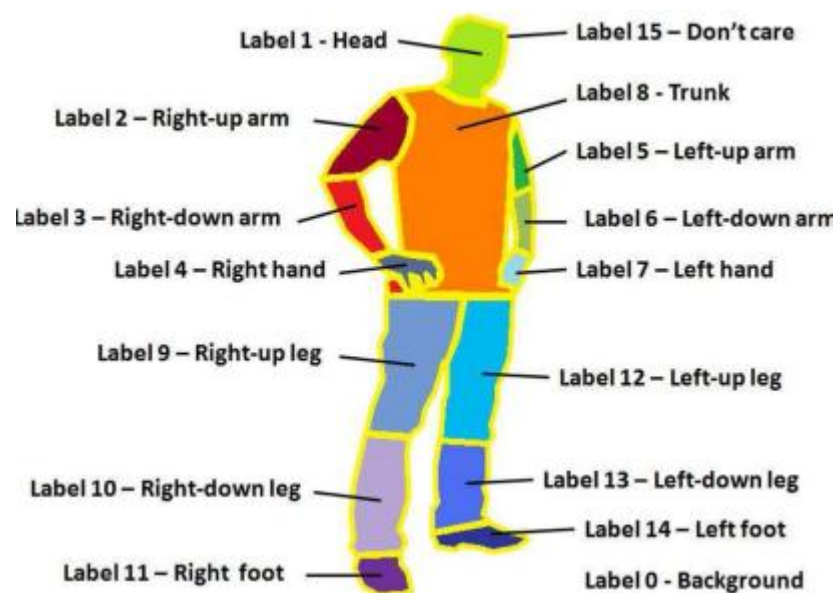


1.2 ADDICIÓ DE MAPES DE PROFUNDITAT PER A L'ANÀLISI D'UN PROBLEMA DE VISIÓ PER COMPUTADOR



- Possibilitat de reconstrucció 3D
- Informació per a identificar el punt de vista de l'àrea d'interès
- Invariant a efectes d'il·luminació i visibilitat
- Evita interferències d'escala (fiabilitat en problemes de fotometria)
- Descripció més acurada a través de RGB-D

1.3 APORTACIONES DELS MAPES DE PROFUNDITAT QUE S'ANALITZEN EN AQUESTA TESI



✓ Desenvolupament d'eines de diagnòstic en el camp clínic basades en fotometria 3D

✓ Models posturals i d'objectes complexos, com les mans, per a l'anàlisi de l'activitat mitjançant descriptors RGB-D



1.4 TECNOLOGIA UTILITZADA



(a)



(b)

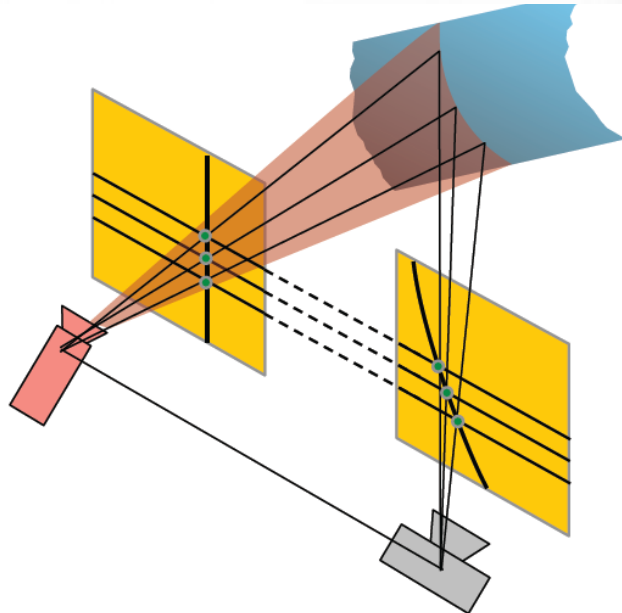


(c)

- Mercat emergent de dispositius RGB-D per la indústria del videojoc.
- *Low-cost.*
- Fàcilment instal·lables.
- Sistema multisensor sincronitzat.



1.4.1 DISPOSITIUS BASATS EN LLUM ESTRUCTURADA



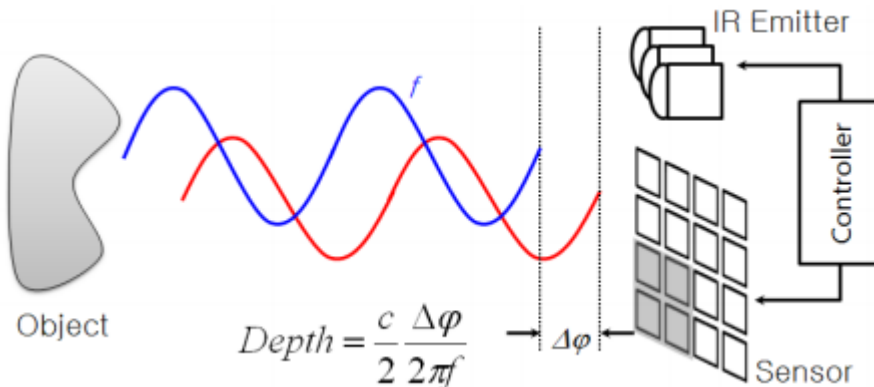
- Sistema format per un projector IR i una càmera sensora
- Baixa resolució
- Vàlid per a rangs curts
- No tolera escenaris exteriors



1.4.2 DISPOSITIUS *TIME OF FLIGHT*



- Mesura el temps en arribar l'ona electromagnètica al sensor per a mesurar la profunditat.
- Alta resolució.
- Tolera certs escenaris exteriors.
- Comportament anòmal en la reflexió d'algunes superfícies
- Vàlid per a rangs curts



OUTLINE

1. Introducció i motivació
2. Reconeixement de la postura basat en descriptors RGB-D
3. Reconeixement de l'activitat basat en mapes de profunditat
4. Descriptor *Spherical Blurred Shape Model*
5. Conclusions
6. Contribucions



2. RECONeixEMENT DE LA POSTURA BASAT EN DESCRIPTORS RGB-D

- Es proposa un disseny d'una metodologia per a la segmentació i discriminació de les diferents parts del cos humà, capaç de descriure la postura de forma discriminant i robusta.
- Disseny i desenvolupament d'una eina d'avaluació postural per al suport al diagnòstic clínic.
- Estudi i validació del seu ús en un entorn clínic.

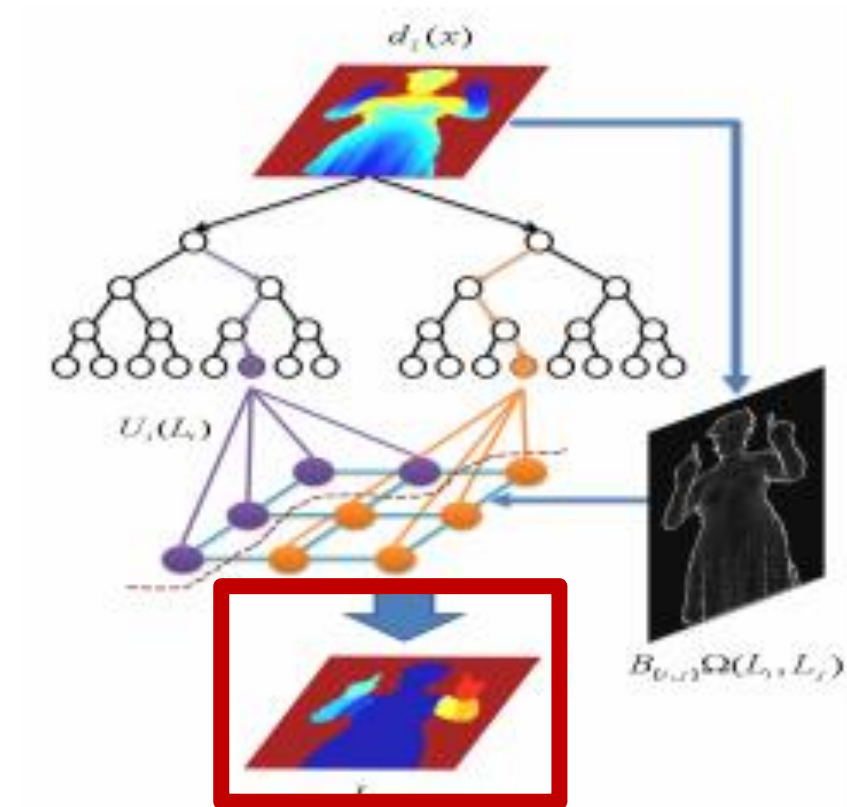


2.1. MÈTODE DE SEGMENTACIÓ EN REGIONS DEL COS HUMÀ

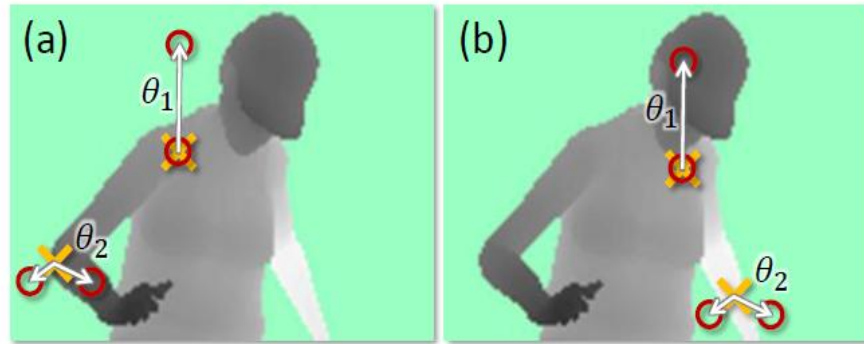
1. Desenvolupament d'un mètode de segmentació de les diferents parts del cos associant un valor de probabilitat de classe al mapa de profunditat. Segmentació de l'escena basada en *Random Forrest*.
2. *Clustering* de probabilitats i classificació final de regió basada en l'algorisme $\alpha - \beta$ *swap Graph-cut*.



2.1. MÈTODE DE SEGMENTACIÓ EN REGIONS DEL COS HUMÀ – WORKFLOW



2.2. CLASSIFICACIÓ DE PÍXELS CANDIDATS A REGIÓ BASADA EN *RANDOM FORREST*



$$f_{\theta}(I, \mathbf{x}) = d_I \left(\mathbf{x} + \frac{\mathbf{u}}{d_I(\mathbf{x})} \right) - d_I \left(\mathbf{x} + \frac{\mathbf{v}}{d_I(\mathbf{x})} \right)$$

$$g_{\theta}(I, \mathbf{x}) = \angle \left(\nabla I \left(\mathbf{x} + \frac{\mathbf{u}}{d_I(\mathbf{x})} \right) - \nabla I \left(\mathbf{x} + \frac{\mathbf{v}}{d_I(\mathbf{x})} \right) \right)$$

$\mathbf{x}$: Píxel a classificar

I : Imatge

$\mathbf{U}, \mathbf{V}$: offsets

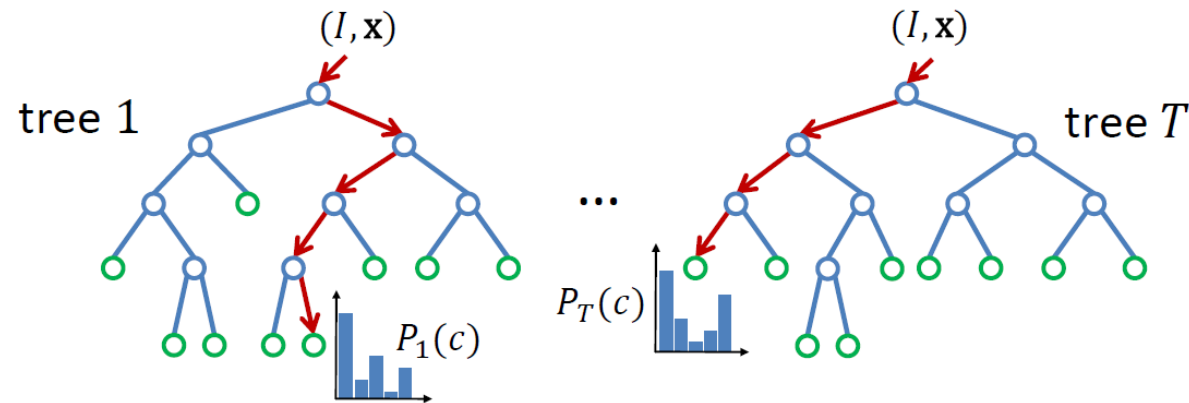
$d_I(\mathbf{x})$: profunditat al píxel $\mathbf{x}$

$\nabla I(\mathbf{x})$: gradient del vector
al píxel $\mathbf{x}$.

- Creació d'un descriptor dels píxels del mapa de profunditat
- Normalització dels *offsets* per a dotar d'invariabilitat respecte la profunditat
- Descriptor invariable translacions 3D



2.2. CLASSIFICACIÓ DE PÍXELS CANDIDATS A REGIÓ BASADA EN *RANDOM FORREST*



- 7 classes (Torso, Braç E, Braç D, Avantbraç E, Avantbraç D, Mà E, Mà D).
- Número màxim d'arbres (3) i profunditat dels arbres (20)
- 500 imatges de *training*

- Objectiu de trobar la combinació d'arbres predictors de classe de manera que cada arbre depengui dels valors d'un vector aleatori provat de forma independent.



2.2.1 RESULTATS DE LA PRIMERA FASE DEL MÈTODE.

- Test format per 100 imatges
- 100 candidats a descriptor *offset*

	Torso	LU arm	LW arm	L hand	RU arm	RW arm	R hand	Avg. per class
100 f_{θ} , $O_{max}=30$, $dt=20$	92.90	73.29	71.42	57.75	74.25	76.26	59.38	72.18
100 f_{θ} , $O_{max}=60$, $dt=20$	94.17	79.83	77.69	77.10	81.04	82.65	80.17	81.81
80 f_{θ} , $O_{max}=60$, $dt=20$	94.22	79.08	76.46	74.19	81.24	83.26	79.05	81.07
60 f_{θ} , $O_{max}=60$, $dt=20$	94.09	78.86	75.86	73.49	79.43	82.60	78.08	80.34
100 f_{θ} , $O_{max}=60$, $dt=15$	94.06	79.81	78.69	76.59	81.18	83.10	80.23	81.95
100 f_{θ} , $O_{max}=60$, $dt=10$	91.83	81.47	78.98	72.30	83.00	83.74	76.85	81.17
60 $f_{\theta}+20 g_{\theta}$, $O_{max}=60$, $dt=20$	94.04	77.73	74.93	71.97	77.62	81.22	76.64	79.17



2.2 OBTENCIÓ DE LES REGIONS DEL COS UTILITZANT *GRAPHCUTS*

1. Donats un conjunt d'imatges $I = \{I^1, \dots, I^s, \dots, I^S\}$ d'una seqüència de vídeo formada per un conjunt de píxels $x = \{x_1, \dots, x_i, \dots, x_{|p|}\}$, volem obtenir un correspondència de regió única a la imatge I
2. La metodologia defineix una funció d'energia $E(L)$ que combina la informació local i contextual. Es basa en la teoria de grafs, on el problema de minimització d'energia basat en el teorema de flux màxim de tall mínim.

$$E(L) = U(L) + \lambda B(L)$$

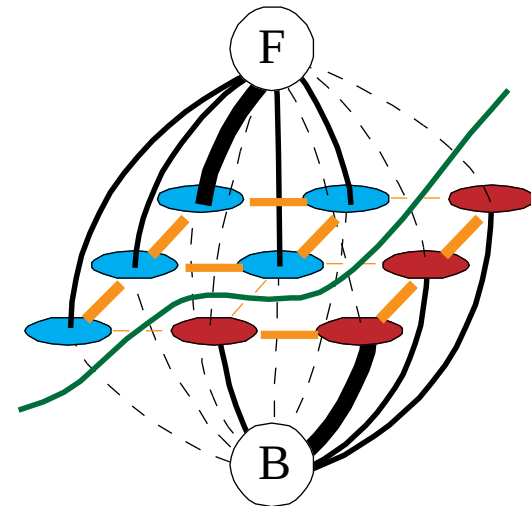


2.2 OBTENCIÓ DE LES REGIONS DEL COS UTILITZANT *GRAPHCUTS*

El potencial unitari conté la probabilitat local de cada píxel de pertànyer a la classe L_i . En aquest cas, utilitzem el logaritme de similitud (*log-likelihood*) de les probabilitats obtingudes per RF.

En el cas del potencial de frontera, utilitzarem la distància entre píxels donat per la següent expressió:

$$B_{\{i,j\}} = \frac{1}{\text{dist}(i,j)} e^{-\beta \|\mathbf{x}_i - \mathbf{x}_j\|^2}$$
$$\beta = (2\langle (\mathbf{x}_i - \mathbf{x}_j)^2 \rangle)^{-1}$$



2.3 RESULTATS FINALS DEL *FRAMEWORK*

Table 2.3: Average per class accuracy in% obtained when applying the different GC approaches –TC: Temporally coherent, Fbf: Frame-by-Frame– , and the best results from the RF approach, in the first row.

	Torso	LU arm	LW arm	L hand	RU arm	RW arm	R hand	Avg. per class
RF results	94.06	79.81	78.69	76.59	81.18	83.10	80.23	81.95
TC, Depth, $\Omega_2 (L_i, L_j)$	98.44	78.93	84.38	88.32	82.57	88.85	93.86	87.91
Fbf, Depth, $\Omega_1 (L_i, L_j)$	98.86	75.05	82.87	91.45	77.57	87.35	93.96	86.73
Fbf, Depth, $\Omega_2 (L_i, L_j)$	98.86	75.03	83.36	92.41	77.54	87.67	94.20	87.01
Fbf, RGB+Depth, $\Omega_1 (L_i, L_j)$	99.02	72.02	81.86	90.29	76.56	86.84	92.14	85.53
Fbf, RGB+Depth, $\Omega_2 (L_i, L_j)$	99.02	72.03	81.95	91.19	76.53	87.12	92.12	85.71

$$\Omega_1(L_i, L_j) = \begin{cases} 0 & \text{for } L_i = L_j \\ 1 & \text{for } L_i \neq L_j \end{cases} \quad \Omega_2(L_i, L_j) = \begin{cases} 0 & \text{for } L_i = L_j \\ 10 & \text{for } L_i = \text{LU}, L_j = \text{RU} \\ & L_i = \text{LH}, L_j = \text{RH} \\ 5 & \text{for } L_i = \text{LW}, L_j = \text{RH} \\ & L_i = \text{RW}, L_j = \text{LH} \\ 1 & \text{otherwise} \end{cases}$$



2.3 RESULTATS FINALS DEL *FRAMEWORK*

	Training images	Average accuracy
J. Shotton (state of the art at 2011) ¹	100k	82,6%
Proposed framework	500	87,91%

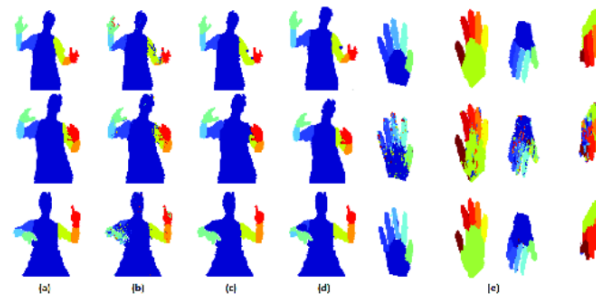


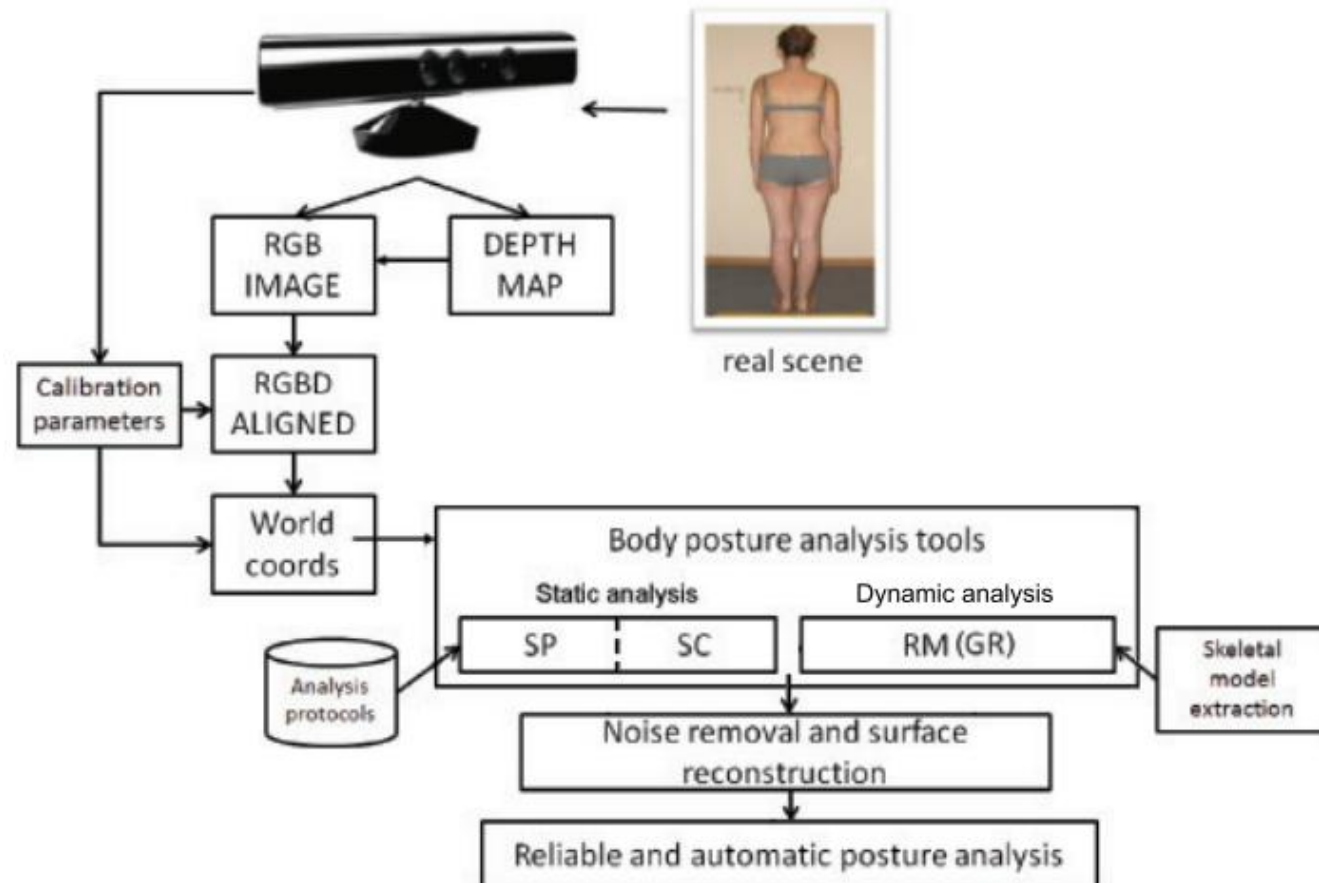
Figure 2.3: Qualitative results; Ground Truth (a), RF inferred results (b), frame-by-frame GC results (c), and Temporally-coherent GC results (d). (e) Hand segmentation experiment. First row shows the ground-truth for two examples. Second row shows the RF classification results. Third row shows the final α -expansion GC segmentation results.

[1] Shotton, A Fitzgibbon, M. C. T. S. M. F. R. M. A. K., and Blake, A. Realtime human pose recognition in parts from single depth images. Computer Vision and Pattern Recognition, IEEE Conference on, Colorado, CO, 2011 (2011).



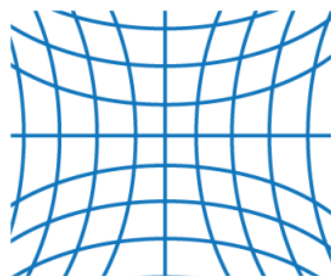
2.4 PROPOSTA D'APLICACIÓ BASADA EN EL MATEIX FRAMEWORK

- Diagrama de funcionament del sistema:

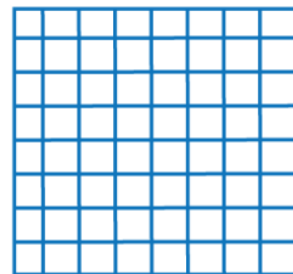


2.4.1 MÒDUL D'OPTIMITZACIÓ DEL MAPA DE PROFUNDITAT

- La informació proveïda pel dispositiu sense ser processada és molt sorollosa i no apta per a realitzar mesures acurades.
- És necessari realitzar un reajustament de les imatges RGB i de profunditat per evitar les següents distorsions:



negative radial distortion
"pincushion"



no distortion



positive radial distortion
"barrel"



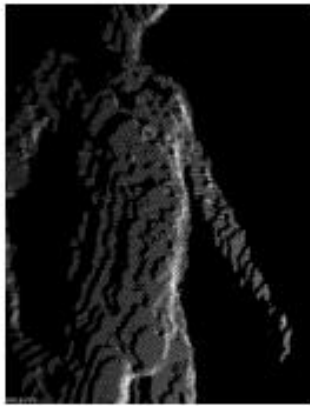
2.4.1 MÒDUL D'OPTIMITZACIÓ DEL MAPA DE PROFUNDITAT

- La resolució dels mapes de profunditat d'aquests dispositius *low-cost* són molt baixes. És necessari realitzar un *upsampling*.
- Depenent de les condicions ambientals de l'escena poden aparèixer regions buides.
- Per evitar aquests problemes realitzem un *resampling* de l'escena basat en l'aplicació de l'algorisme *Radial Basis Function*¹

J.C. Carr and R.K. Beatson, Reconstruction and Representation of 3D Objects with Radial Basis Functions, Siggraph 2003.



2.4.1 MÒDUL D'OPTIMITZACIÓ DEL MAPA DE PROFUNDITAT



(a)



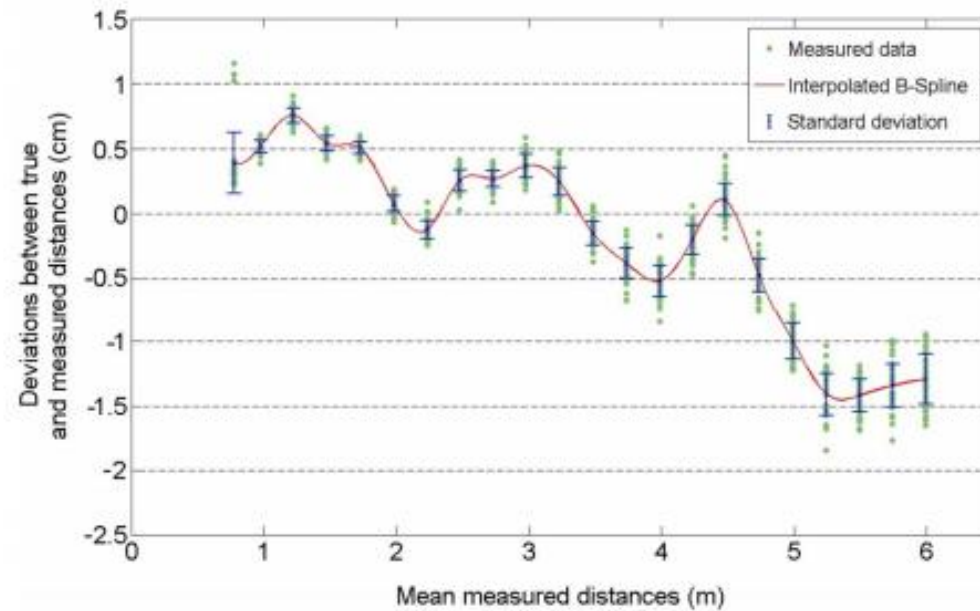
(b)



(c)



(d)

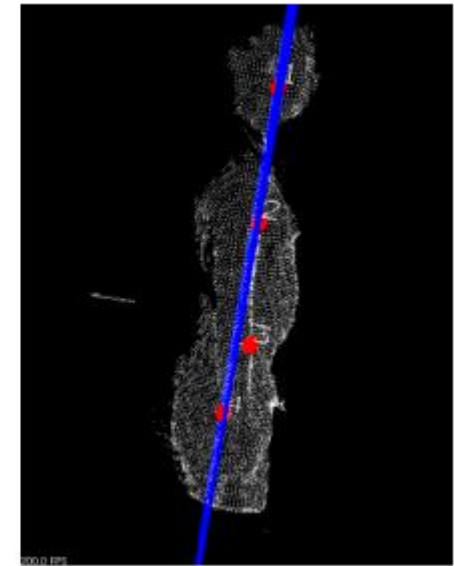
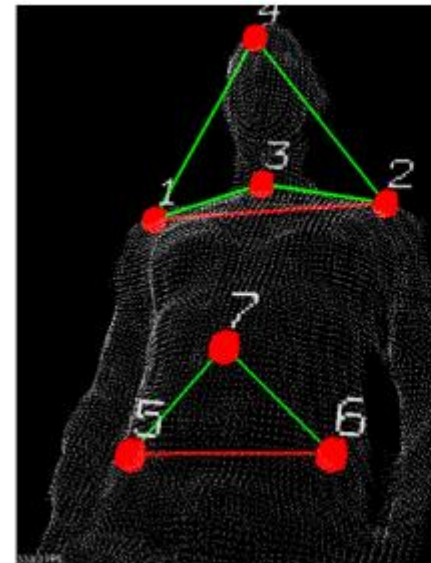
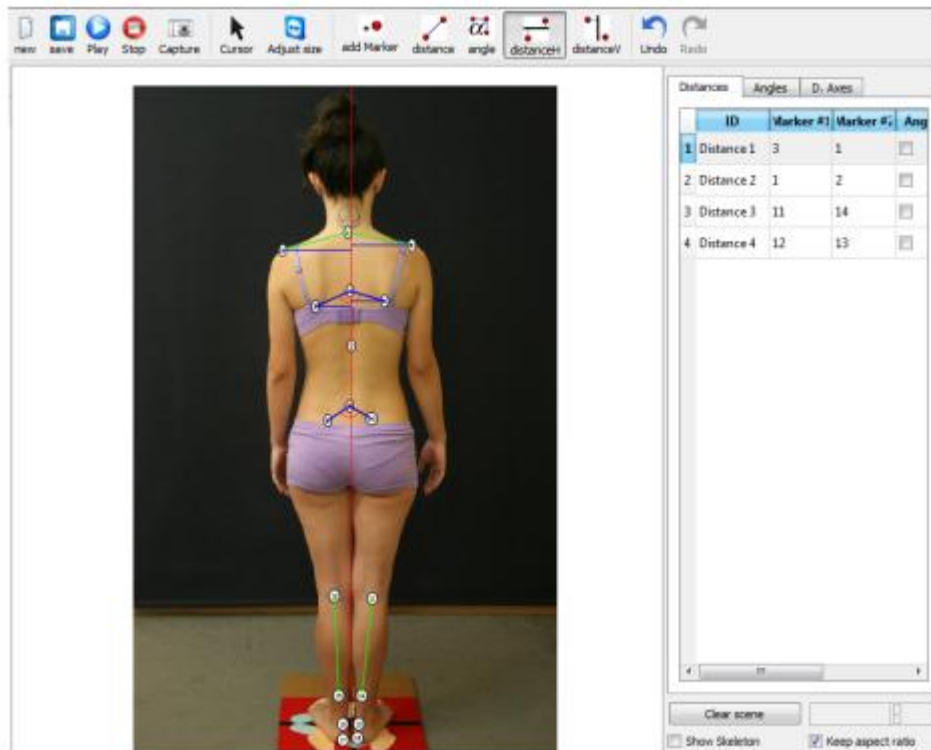


(a)

✓ El sistema té un bon comportament a entre 1,8 i 2,2 metres de distància respecte l'objecte a analitzar



2.4.2 ANÀLISI ESTÀTIC DE LA POSTURA



✓El sistema detecta regions on aplicar el protocol d'anàlisi postural



2.4.3 ANÀLISI ESPINAL

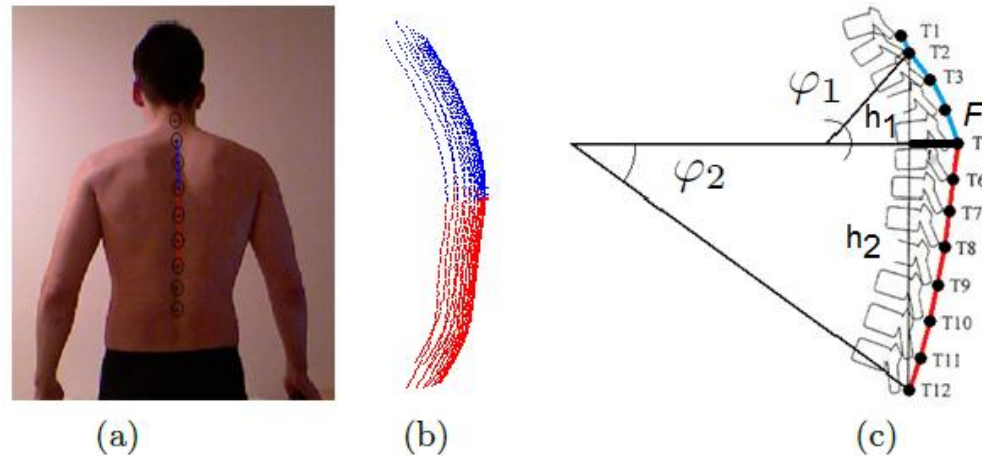


Figure 2.10: (a) Sample of analysis. (b) Automatically reconstructed 3D spinal cloud. (c) Geometric model to obtain anthropometric kyphosis and lordosis value.

- ✓ El sistema obté coordenades 3D dels punts de referència per tal d'obtenir els angles de lordosi i cifosi basat en el treball de Leroux¹

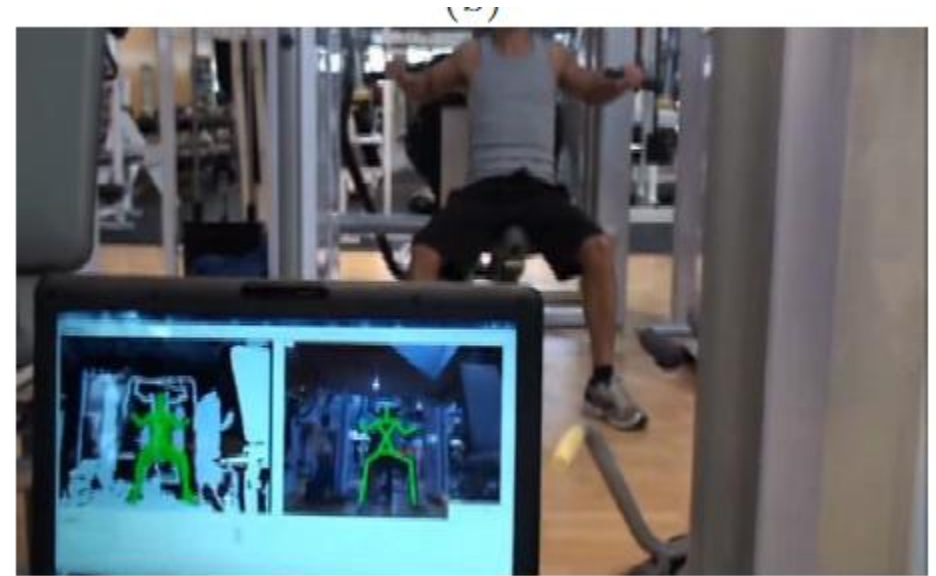
[1] Leroux, M. A., and Zabjek, K. A noninvasive anthropometric technique for measuring kyphosis and lordosis: application for scoliosis. Spine, Pubmed, Université de Montréal and Sainte-Justine Hospital, Montréal



2.4.4 ANÀLISI DINÀMIC



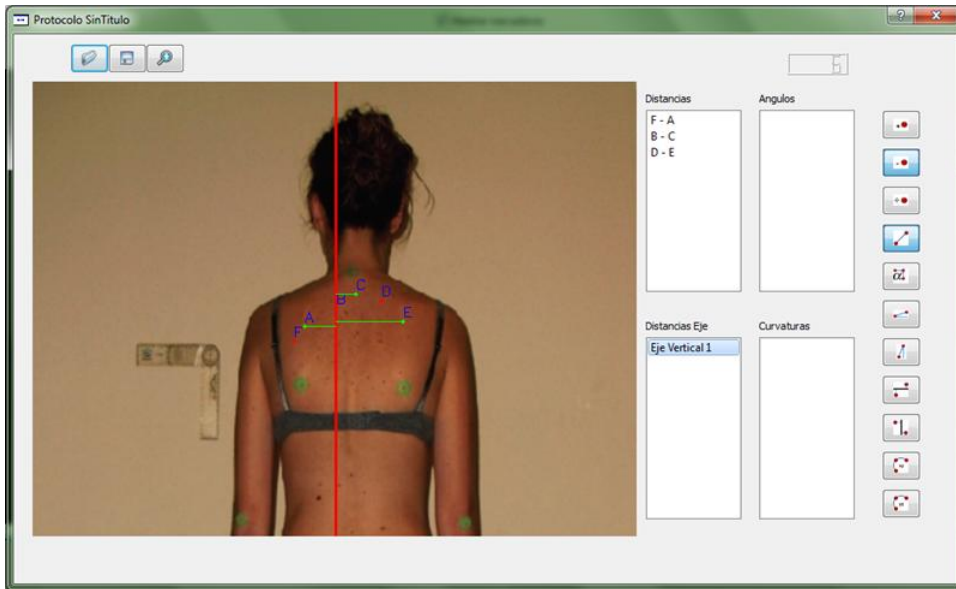
Figure 2.13: Skeletal model and example of selected articulations with computed dynamic range of movement (maximum opening and minimum closing values of a joint measured in degrees for a certain period of time).



- ✓ Anàlisi del rang de moviment basat en el seguiment dels marcadors 3D.



2.4.5 VALIDACIÓ DEL SISTEMA



- ✓ Validació clínica realitzada amb el suport de l'Institut de Fisioteràpia Global Mézières.
- ✓ Comprovació dels resultats amb eines homologades i comparades amb 3 observadors independents.
- ✓ Validació exhaustiva de múltiples mesures de regions corporals en diferents configuracions.



2.4.6 RESULTATS DEL SISTEMA

Distance subject-device (m)	1,3	1,9	2,2
AAV (° movement)	2,2	3,8	5,2
AAV (mm)	0,98	1,42	2,1
AAV (° angles)	0,51	1,04	1,24
AAV (%)	0,46	0,77	1,3
Standard Error (%)	1,01	1,18	1,71

Table 2.4: Pose and range of movement precision.

	Khyposis range	Lordosis range
AAV (°)	5	6

Table 2.5: Validation of spinal analysis.

- ✓ Els resultats finals fan factible aquesta eina com a sistema de suport al diagnòstic clínic de transtorns músculo-esquelètics.
- ✓ Dades de 2011 emprant un dispositiu de llum estructurada. Actualment s'utilitza un sistema ToF que millora un 40% els resultats.



OUTLINE

1. Introducció i motivació
2. Reconeixement de la postura basat en descriptors RGB-D
3. Reconeixement de l'activitat basat en mapes de profunditat
4. Descriptor *Spherical Blurred Shape Model*
5. Conclusions
6. Contribucions



3. RECONeixEMENT DE L'ACTIVITAT BASAT EN MAPES DE PROFUNDITAT

- Es proposa el disseny d'una metodologia per a la detecció, en temps real, d'activitats i gestos a través de l'anàlisi multimodal de les dades extretes d'un dispositiu RGB-D.
- Classificació multiclasse d'un conjunt d'activitats i gestos, capaç de descriure un comportament.



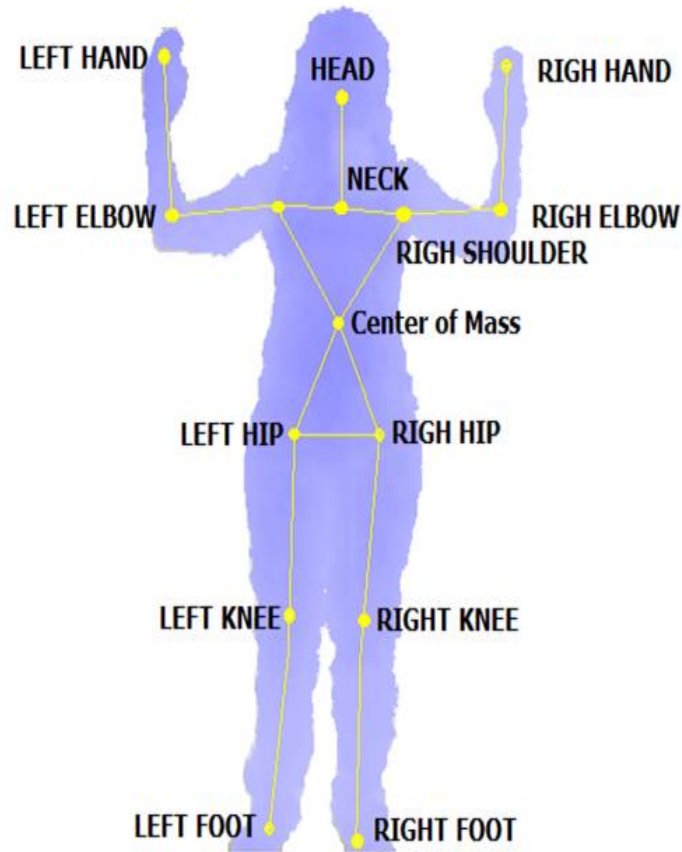
3. RECONeixEMENT DE L'ACTIVITAT BASAT EN MAPES DE PROFUNDITAT



- Mapa de profunditat directament extret del dispositiu de captura *low-cost* rgb-d
- Distinció *persona/background* basada en RF només emprant mapa de profunditat.
- Obtenció de les regions corporals basades en *mean-shift* amb l'objectiu de treball en temps real.



3.1. DESCRIPTOR DE L'ACTIVITAT BASAT EN MARCADORS CORPORALS

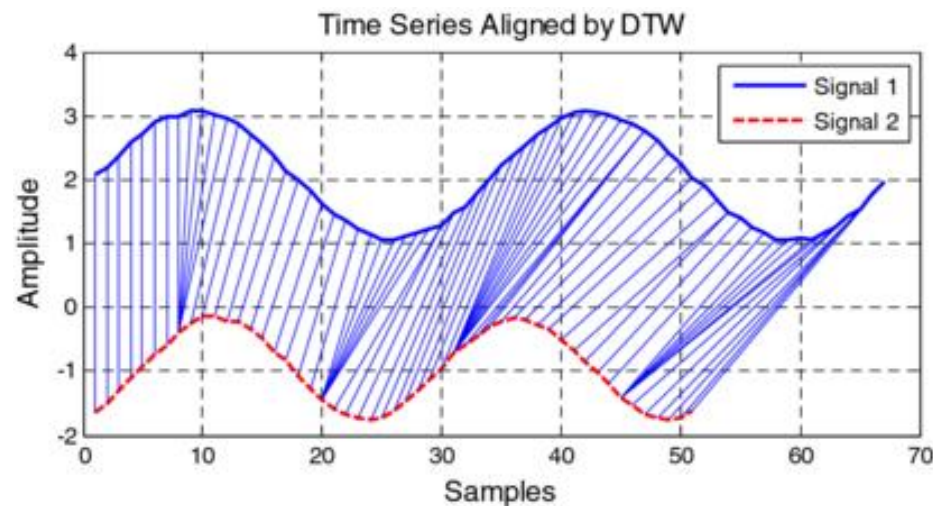


- Obtenció de 14 punts corporals tridimensionals.
- El vector de característiques que descriu la posició corporal a l'instant j ve donat per:

$$V_j = \{\{v_{j,x}^1, v_{j,y}^1, v_{j,z}^1\}, \dots, \{v_{j,x}^{14}, v_{j,y}^{14}, v_{j,z}^{14}\}\}$$



3.2. CLASSIFICACIÓ DE L'ACTIVITAT BASADA EN *DYNAMIC TIME WARPING*



- DTW és capaç de realitzar una mesura de similaritat entre dues sèries temporals amb variabilitat en el temps



3.3. FEATURE WEIGHTING IN DTW

Input: Ground-truth data formed by N sets of gestures $\{n_1, \dots, n_N\}$.
Output: Weight vector $\nu = \{\nu^1, \dots, \nu^z\}$ associated with skeletal joints so that $\sum_{i=1}^z \nu^i = 1$.
 $\nu = \emptyset$
for $p = 1 : z$ do // Number of joints
 for $i = 1 : N$ do
 for $j = i : N$ do
 $D^p(i, j) = \text{mean}(\text{DTW}(C_v^i, C_w^j)), \forall v, w$
 gesture samples of categories i and j .
 end
 end
 $\nu_{\text{intra}} = \frac{\sum_{i=1}^{N-1} \sum_{j=i+1}^N D^p(i, j)}{\frac{N \times (N-1)}{2}}$ // Computer intra-class variability
 $\nu_{\text{inter}} = \frac{\text{Trace}(D^p)}{m}$ // Computer inter-class variability
 $\nu^p = \max(0, \frac{\nu_{\text{intra}} - \nu_{\text{inter}}}{\nu_{\text{intra}}})$ // Compute global weight for joint p
 $\nu = \nu \cup \nu^p$
end
Normalize ν so that $\sum_{i=1}^z \nu^i = 1$

- Es proposa una millora a l'algorisme de detecció de patrons DTW basat en l'addició de pesos als elements del vector de característiques.
- Construcció d'un dataset format
- Maximització dels coeficients intraclasses i interclasses per a una major discriminació de les seqüències.
- Obtenció del *threshold* mínim per a discriminar una seqüència.



3.4. FEATURE WEIGHTING IN DTW

Gesture training database

G1

G2

...

Gn

G1₁

G2₁

Gn₁

G1₂

G2₂

Gn₂

...

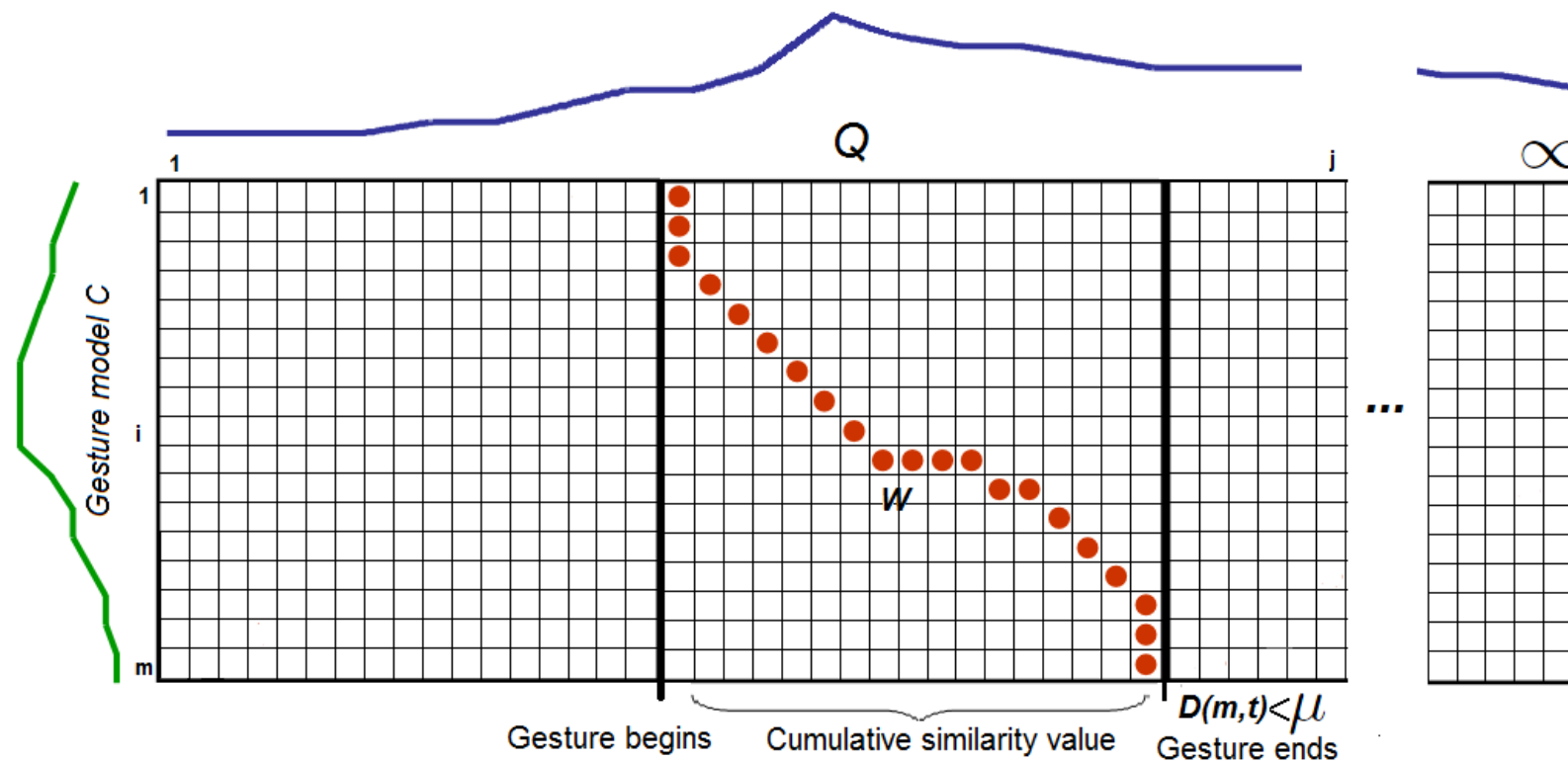
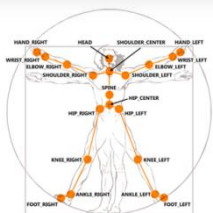
...

...

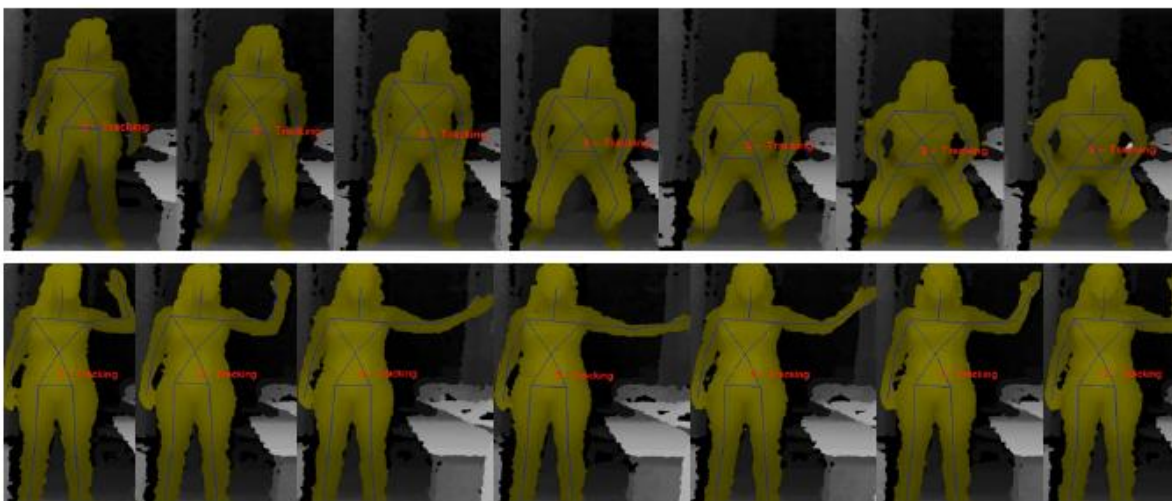
G1₃

G2₃

Gn₃



3.5. RESULTATS DE *FEATURE WEIGHTING* IN DTW



Classification Results Feature Weighting DTW		
Gesture	Begin-end DTW	Feature Weighting
Jump	68	68
Bend	63.4	68
Clap	42	55
Greet	64.2	73
Note	68	76

- La millora del DTW no es produeix de forma homogènia
- Altament depenent de la naturalesa de l'activitat



3.6. ORGANITZACIÓ CHALEARN



(1) Vattene (2) Viene qui (3) Perfetto (4) È un furbo (5) Che due palle



(6) Che vuoi (7) d'accordo Vanno (8) Sei pazzo (9) Cos hai combinato (10) Non me frrega niente



(11) Ok (12) Cosa ti farei (13) Basta (14) Le vuoi prendere (15) Non ce ne più



(16) Ho fame (17) Tanto tempo fa (18) Buonissimo (19) Si sono messi d'accordo (20) Sono stufo

- Organització d'una competició, a nivell mundial, de discriminació de seqüències basada en informació multimodal.
- 20 categories relacionades amb gestos típics de la cultura italiana.
- Dades: RGB, mapa profunditat, model esquelètic i àudio.
- 11116 seqüències (1-2 min) de *training/testing*
- 80% Italians nadius i 20% Italoparlants



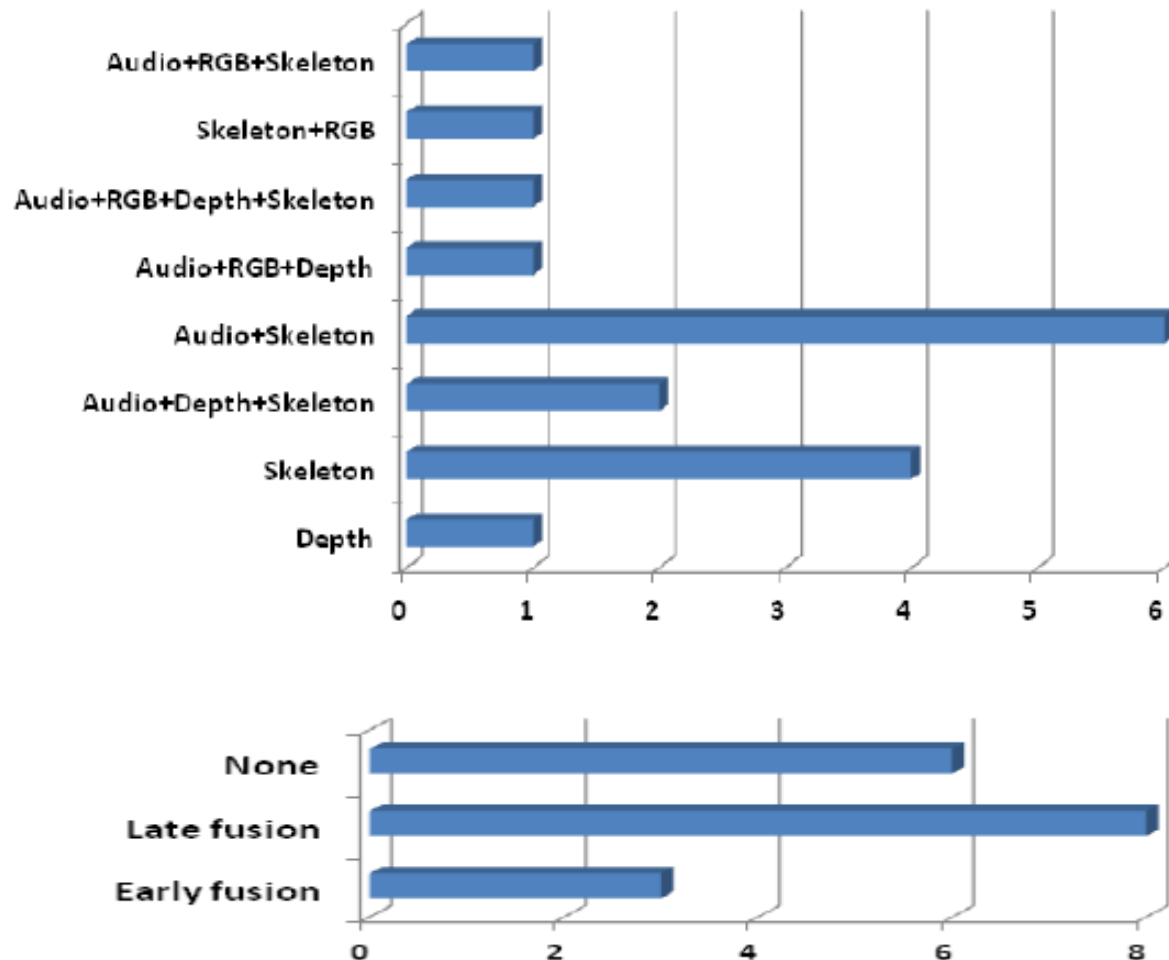
3.6.1 RESULTATS DE LA COMPETICIÓ

TEAM	Validation score	Test score
IVA MM	0.20137	0.12756
WWEIGHT	0.46163	0.15387
ET	0.33611	0.16813
MmM	0.25996	0.17215
PPTK	0.15199	0.17325
LRS	0.18114	0.17727
MMDL	0.43992	0.24452
TELEPOINTS	0.48543	0.25841
CSI MM	0.32124	0.28911
SUMO	0.49137	0.31652
GURU	0.51844	0.37281
AURINKO	0.31529	0.63304
STEVENWUDI	1.43427	0.74415
JACKSPARROW	0.86050	0.79313
JOEWAN	0.13653	0.83772
MILAN KOVAC	0.87835	0.87463
IAMKHADER	0.93397	0.92069

- A la competició es van presentar 54 equips formats principalment per universitats i centres d'investigació de tot el món.
- La competició es va realitzar en dos fases. 17 equips van realitzar el test final.
- Els competidors van enviar el codi i una explicació del mètode emprat.
- El guanyador final va obtenir un error del 12,75%.



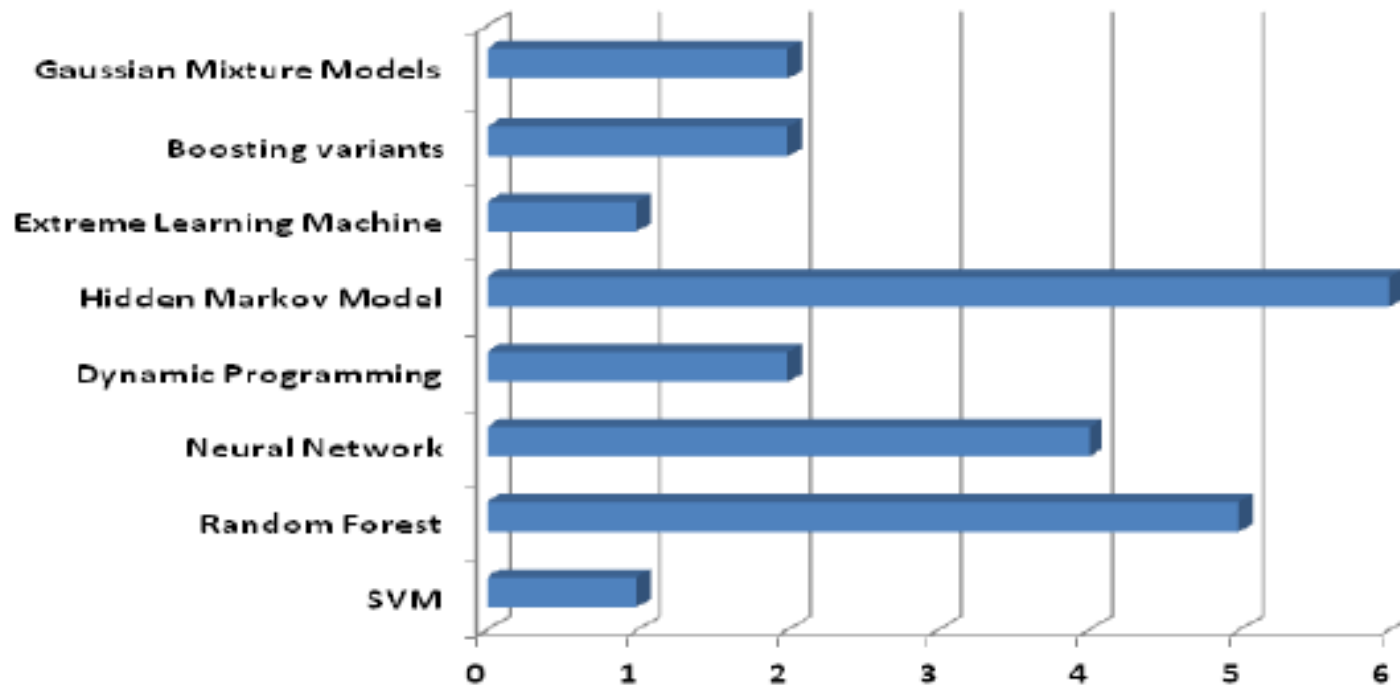
3.6.2 OBSERVACIONS DE LA COMPETICIÓ



- Principalment els canals d'informació més utilitzats són l'àudio i el model esquel·t.
- Cap participant ha utilitzat només l'àudio
- Existeix força variabilitat a l'hora de realitzar una fusió dels vectors de característiques. L'estratègia més utilitzada es tractar de forma independent cada canal i fusionar dins del classificador



3.6.2 OBSERVACIONS DE LA COMPETICIÓ



- Ampli ventall d'opcions de classificació
- En tots els casos s'han utilitzat mètodes supervisats, excepte en un cas.
- Tradicionalment sempre s'ha utilitzat HMM per a la classificació de seqüències. A l'estudi es confirma la tendència.



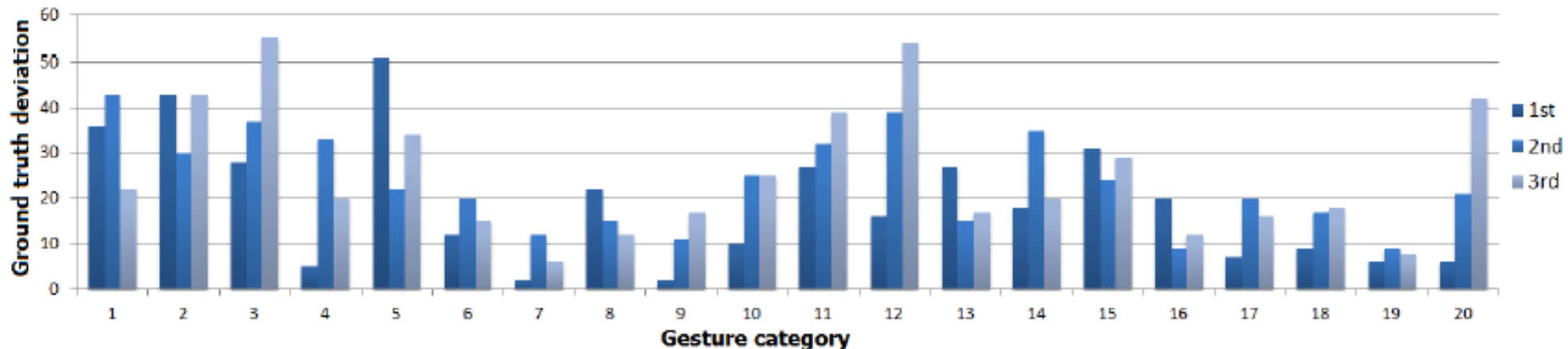
3.6.2 OBSERVACIONS DE LA COMPETICIÓ

- El primer (87,3%), segon (86,6%) y tercer (83,2%) classificats han utilitzat la informació d'àudio i el model esquel·tural.
- El primer classificat utilitza una estratègia de classificació combinada en *Gaussian Hidden Markov Model* i *DTW*
- El segon classificat utilitza una estratègia d'extracció de característiques 3D del model esquel·tural i l'espectrograma d'àudio, com a descriptor d'un *Random Forrest*.
- El tercer classificat combina dos estratègies. En primera instància, realitza un *clustering* no supervisat de les característiques d'àudio per a entrenar un *RF*. En segon terme, generen un conjunt de training, a través d'estadística descriptiva, associant de la sortida del clasificador amb el model esquel·tural.



3.6.2 OBSERVACIONS DE LA COMPETICIÓ

- Els gestos 6, 7, 8, 9, 16, 17, 18 i 19 tenen un comportament molt semblant entre els diferents mètodes. Els gestos 1, 2, 3, 5 i 12 tenen una baixa precisió.
- La informació estructural al voltant del model esquelètic és útil per a la discriminació de categories gestuals que contenen trajectòries similars.
- En competicions posteriors s'observa la emergència de mètodes basats en Deep Learning (CNN), on l'accuracy arriba al 95%.



OUTLINE

1. Introducció i motivació
2. Reconeixement de la postura basat en descriptors RGB-D
3. Reconeixement de l'activitat basat en mapes de profunditat
4. Descriptor *Spherical Blurred Shape Model*
5. Conclusions
6. Contribucions



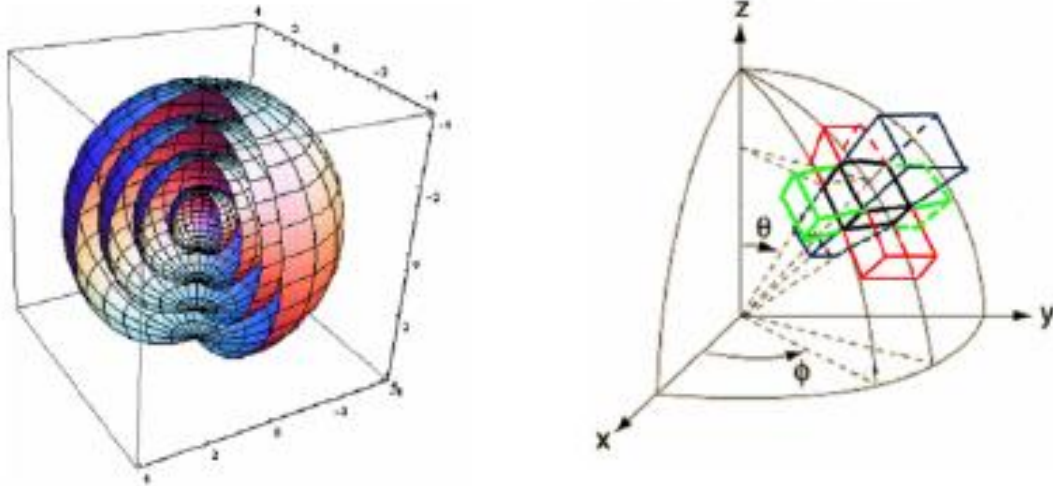
4. DESCRIPTOR *SPHERICAL BLURRED SHAPE MODEL*

- Tot i que ja existeixen descriptors 3D en el camp de la visió per computador encara tenen mancances en quant a discriminació i computacional eficients.
- L'objectiu és dissenyar un descriptor 3D suficientment robust per a elements complexos com poden ser les mans d'una persona, on constantment es troben oclusions parcials.
- És necessari el disseny d'un descriptor invariant a rotacions i escala.
- Es proposa una classificació multiclasse per tal de codificar diferents estats i formar un vocabulari.



4.1.1 CARACTERÍSTIQUES DEL DESCRIPTOR

- L'objectiu és codificar la densitat estructural d'un objecte 3D, juntament, amb les variabilitats locals extrems del veïnatge entre voxels.



$$B = \{b_{\{0,0,0\}}, \dots, b_{\{i,j,k\}}, \dots, b_{\{N_L-1, N_\theta-1, N_\phi-1\}}\},$$

$$\forall i \in \{0, 1, \dots, N_L - 1\}, \forall j \in \{0, 1, \dots, N_\theta - 1\},$$

$$b_{\{i,j,k\}}^* = (i \cdot d_R + \frac{d_R}{2}, j \cdot d_\theta + \frac{d_\theta}{2}, k \cdot d_\phi + \frac{d_\phi}{2}),$$

$$\forall i \in \{0, 1, \dots, N_L - 1\}, \forall j \in \{0, 1, \dots, N_\theta - 1\},$$

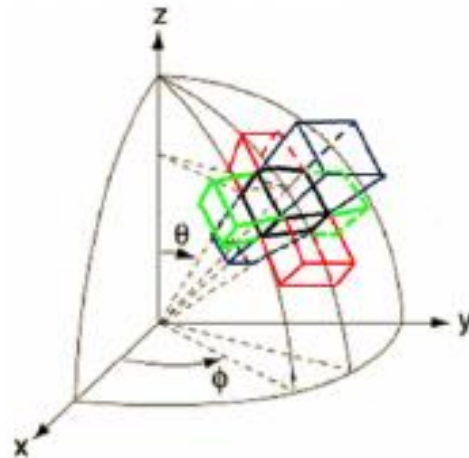
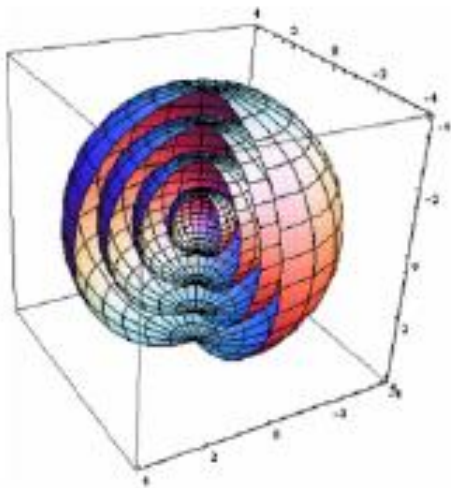
$$\forall k \in \{0, 1, \dots, N_\phi - 1\}.$$

$$P = \{p_i | p_i \in \mathbb{R}^3\}, C, N_L, N_\theta, N_\phi, R, \sigma$$



4.1.1 CARACTERÍSTIQUES DEL DESCRIPTOR

- L'objectiu es codificar la densitat estructural d'un objecte 3D, juntament, amb les variabilitats locals extrems del veïnatge entre voxels. Connectivitat 26+1.



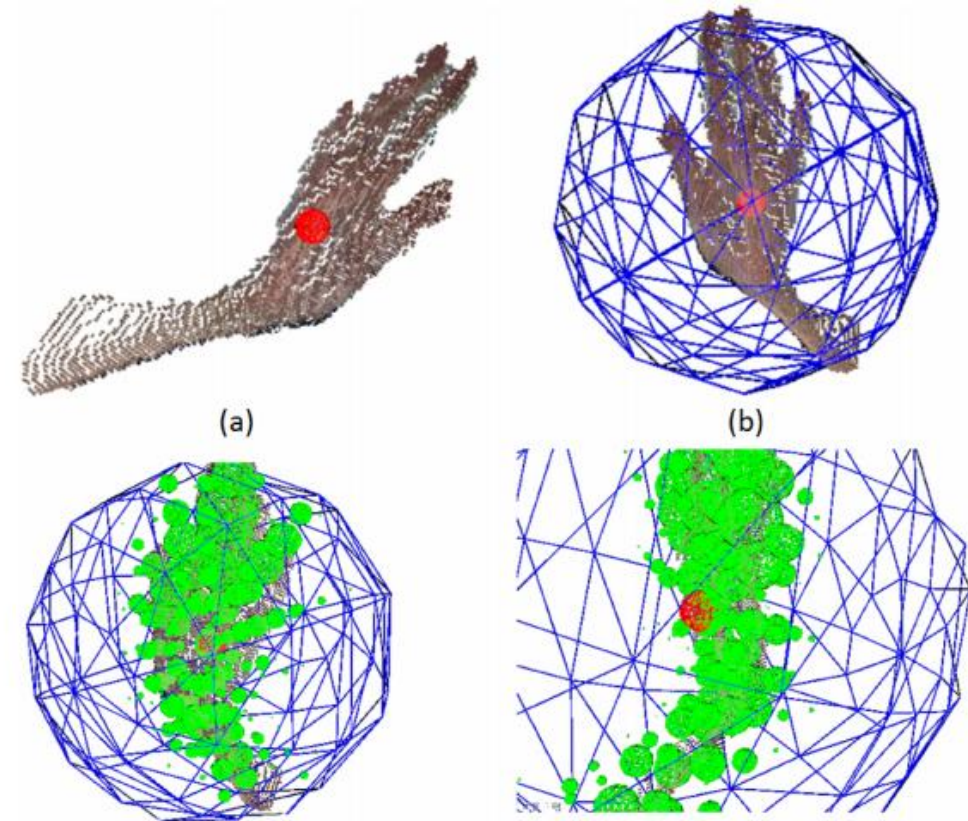
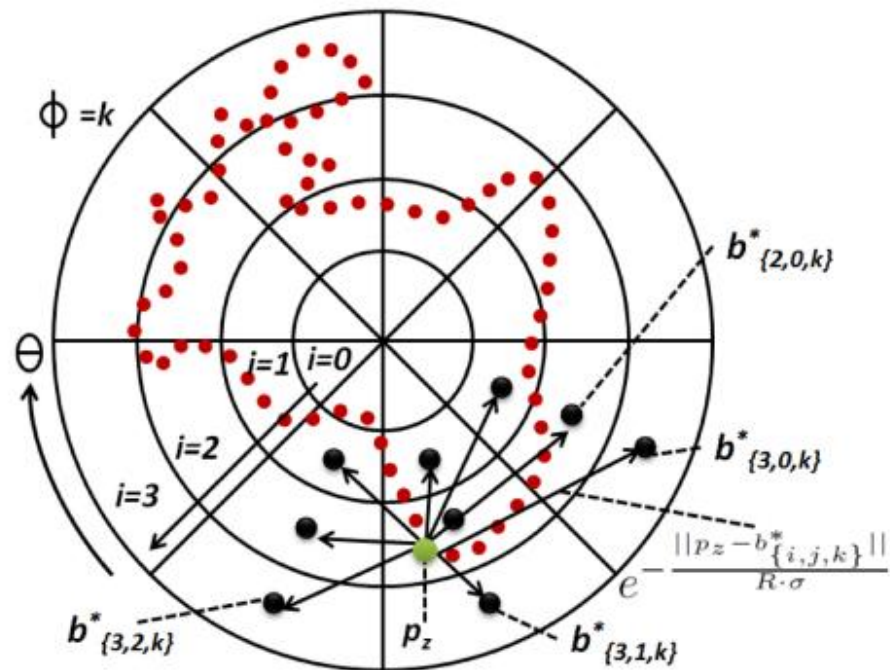
$$W_i = 0, \forall i \in \{1, 2, \dots, N_L \cdot N_\theta \cdot N_\phi\}.$$

$$W^* = W^* \cup \left\{ e^{-\frac{\|p_z - b_{\{i,j,k\}}^*\|}{R \cdot \sigma}} \right\}$$

$$W_i = \frac{W_i}{\sum_{j=1}^{N_L \cdot N_\theta \cdot N_\phi} W_j}, \forall i \in \{1, 2, \dots, N_L \cdot N_\theta \cdot N_\phi\}.$$

$$P = \{p_i | p_i \in \mathbb{R}^3\}, C, N_L, N_\theta, N_\phi, R, \sigma$$

4.1.2 REPRESENTACIÓ GRÀFICA DEL DESCRIPTOR 3D



4.1.2 CARACTERÍSTIQUES DEL DESCRIPTOR

- L'objectiu és dotar al descriptor de ser invariant al punt de vista 3D, es rota a partir de la informació d'orientació extretat de l'anàlisi de cada un dels *bins*.
- Utilitzarem la orientació de l'eix de major densitat com a part del vector de característiques.

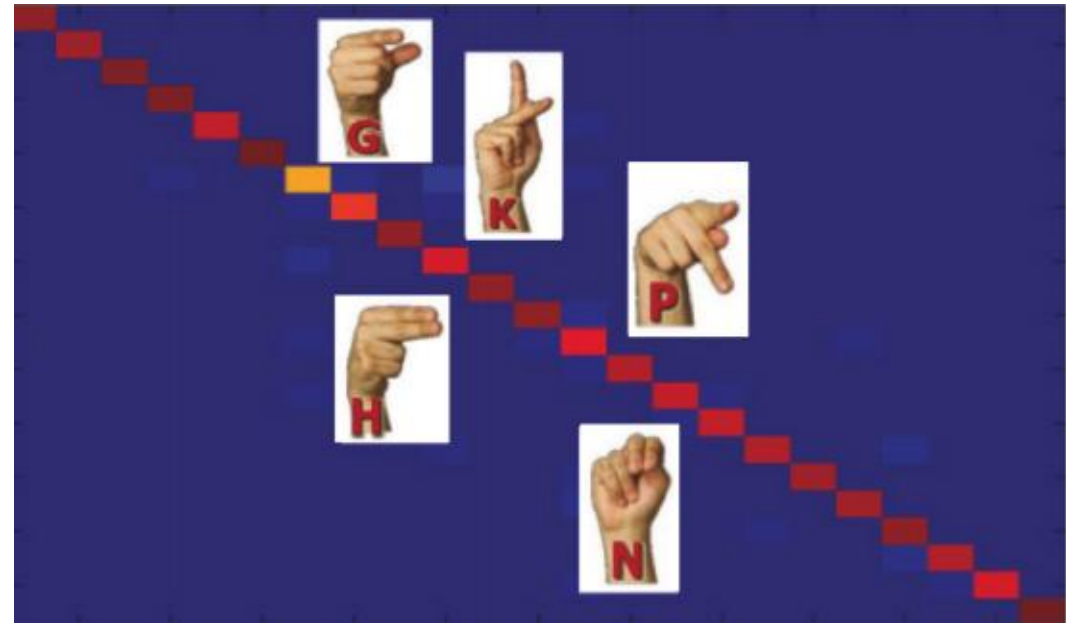
$$f(\theta, \phi) = \sum_{r=1}^{N_L} W_{\{r, \theta, \phi\}}, \quad \vec{T}_1 = \arg \max_{\theta, \phi} f(\theta, \phi), \quad \vec{T}_2 = \arg \max_{\theta, \phi \setminus \vec{T}_1} f(\theta, \phi).$$

- La orientació és representada en forma de quaternions amb l'objectiu d'evitar ambigüetats.



4.2 VALIDACIÓ DEL SISTEMA

- Amb l'objectiu d'avaluar el rendiment del descriptor 3D s'escullen dos *datasets RGB-D* públics.



4.2.1 DETALLS DE LA VALIDACIÓ DEL SISTEMA

- El classificador utilitzat es *Support Vector Machine*
- Estratègia multiclasse *one vs one* i *kernel RBF*
- El número de capes i subdivisions angulars és 8 i radi 0,15m
- Principalment comparem el seu rendiment amb VFH¹ i ESF², degut a que són la referència actual.

[1] Rusu, R., Bradski, G., Thibaux, R., and Hsu, J. Fast 3d recognition and pose using the viewpoint feature histogram. International Conference on Intelligent Robots and Systems (IROS) (2010), 2155{2162}.

[2] Wohlkinger, W., and Vincze, M. Ensemble of shape functions for 3d object classsication. IEEE Robotics and Biomimetics (2011).



4.2.2 RESULTATS DE LA VALIDACIÓ DEL SISTEMA

- RGB-D Objects dataset

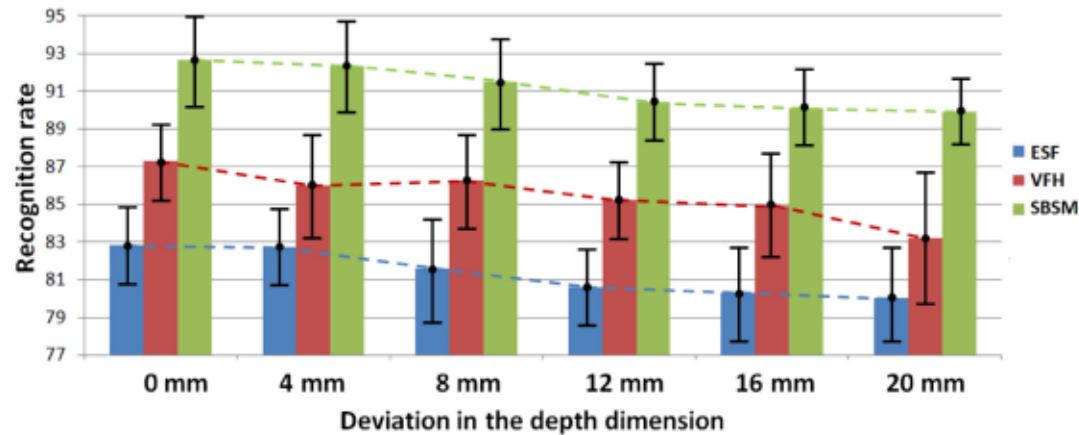
Method	Depth performance
SIFT + Texton + Color + Spin [49]	64.7%
Sparse Distance Learning [50]	70.2%
VFH + SVM	77.5%
RGB-D Kernel Descriptors [12]	80.3%
Hierarchical Matching Pursuit [13]	81.2%
ESF + SVM	84.9%
SBSM, $\sigma = 0$ + SVM	96.7%
SBSM, $\sigma = 1$ + SVM	97.9%

- ASL dataset

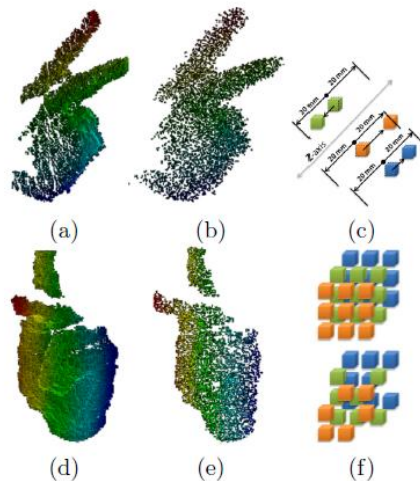
Method	Depth performance
VFH + SVM	88.7%±0.2
ESF + SVM	92.3%±0.2
SBSM, $\sigma = 0$ + SVM	97.1%±0.6
SBSM, $\sigma = 1$ + SVM	99.3%±0.4



4.2.2 RESULTATS DE LA VALIDACIÓ DEL SISTEMA



- Robust davant transformacions de rotació i escala
- Robust davant baixes resolucions
- Robust davant soroll produït per fortes acceleracions



4.3 ANÀLISI QUALITATIU DEL *SBSM*

- Es proposa un disseny i desenvolupament d'un sistema basat en el reconeixement i classificació d'objectes 3D
- El sistema ha de poder funcionar en temps real
- Es pot combinar amb interacció gestual



4.3 ANÀLISIS QUALITATIU DEL SBSM

**Spherical Blurred Shape
Model Descriptor for
3D Object Recognition**



OUTLINE

1. Introducció i motivació
2. Reconeixement de la postura basat en descriptors RGB-D
3. Reconeixement de l'activitat basat en mapes de profunditat
4. Descriptor *Spherical Blurred Shape Model*
5. **Conclusions**
6. Contribucions



5. CONCLUSIONS. FRAMEWORK SEGMENTACIÓ

- ✓ Es proposa un framework genèric per a la segmentació del cos basat en *Random Forrest* i *Graph-cut* amb el benefici de la coherència espacio-temporal.
- ✓ El classificador *Random Forest* és capaç d'estimar la probabilitat de cada píxel del mapa de profunditat i associar a una regió del cos.
- ✓ *Graph-cuts* és utilitzat per optimitzar la posició espacial i temporal de les probabilitats de classe rebudes per *RF*.
- ✓ Els resultats obtinguts demostren la seva eficàcia per a segmentar vàries parts del cos i demostra superioritat davant d'altres metodologies.



5. CONCLUSIONS. EINA DE SUPORT AL DIAGNÒSTIC CLÍNIC

- ✓ Es proposa un disseny i una implementació per a un sistema d'avaluació de la postura i l'anàlisi espinal, obtinguda a través de dispositius RGB-D *low-cost*.
- ✓ El sistema es compon d'integrar múltiples disciplines del camp de la visió per computador, *machine learning*, anàlisi de sèries temporals i visualització 3D
- ✓ El sistema és capaç de ser una eina de suport per a la prevenció i tractament de desordres músculo-esquelètics. S'inclou l'eina en la rutina clínica en àrees com la reeducació postural, rehabilitació i la salut esportiva

5. CONCLUSIONS. FRAMEWORK RECONeixEMENT GESTOS

- ✓ Es proposa un sistema totalment automàtic per al reconeixement d'activitats basat en mapes de profunditat en escenaris no controlats.
- ✓ El sistema és capaç d'analitzar seqüències de dades basades en l'assignació de pesos als vectors de característiques per a que un *DTW* els pugui discriminar de forma eficaç.
- ✓ L'assignació dels pesos es realitza mitjançant l'anàlisi dels coeficients intraclasse e interclasse de les diferents categories d'activitats.
- ✓ L'avaluació del mètode demostra novetat en el reconeixement d'activats en mapes de profunditat millorant parcialment els resultats aplicant mètodes classics de *DTW*.

5. CONCLUSIONS. COMPETICIÓ RECONeixEMENT DE GESTOS

- ✓ Es construeix un *dataset* públic compostat per 20 gestos típics de la cultura italiana, amb informació multimodal.
- ✓ El mètode més utilitzat és *Hidden Markov Models*, com és el cas del guanyador.
- ✓ La informació més utilitzada per els participants és l'àudio i model esquelètic.
- ✓ La informació estructural al voltant de les coordenades del model esquelètic, proporcionen discriminació de categories gestuals que contenen trajectòries similars.

5. CONCLUSIONS. DESCRIPTOR 3D

- ✓ Es presenta el descriptor *Spherical Blurred Shape Model* per a la identificació d' objectes 3D
- ✓ En la validació del mètode els resultats superen en 16,7% l'estat del art
- ✓ El descriptor és invariant a diferents distorsions i transformacions 3D, alhora de computacionalment eficient per aplicacions en temps real.
- ✓ El descriptor és transferit a sistemes *Human Computer Interaction* funcionant de manera robusta.

OUTLINE

1. Introducció i motivació
2. Reconeixement de la postura basat en descriptors RGB-D
3. Reconeixement de l'activitat basat en mapes de profunditat
4. Descriptor *Spherical Blurred Shape Model*
5. Conclusions
6. Contribucions

6. CONTRIBUCIONS

15 Publicacions en congressos nacionals internacionals:

- *Computer Vision / Machine Learning*
- Sistemes i aplicacions
- Àrea clínica

CVPR

ICCV

ICPR

4 Publicacions en revistes d'impacte:

- *Computer Vision / Machine Learning*
- Sistemes i aplicacions

PRL

Transactions
on Cybernetics

Computer in
industry

1 Escriptura d'un capítol llibre:

- *Computer Vision / Machine Learning*

Google Scholar

Citation indices	All	Since 2012
Citations	484	478
h-index	10	10
i10-index	11	11



6. CONTRIBUCIONS

Adscrit a 4 grups d'investigació:

- *HUPBA: Human Processing Behavior Analysis*
- *Applications to the analysis of Medical Imaging and eHealth sector*
- *MEDiminder*
- *BCN Perceptual Group*

Premi Chalearn & Microsoft Competition:

M. Reyes, O. Lopes, M. Pousa, J. Gonzalez, and S. Escalera. Third place award Chalearn Gesture Challenge, Demonstration Competition at International Congress of Pattern Recognition. Organized by Microsoft Research. Tokyo, November 2012

Nominat als premis MIT Innovators under 35 Europe

6. CONTRIBUCIONS

Creació d'una *spin-off*, PhysicalTech SL, conjuntament Universitat de Barcelonai l'Institut de Fisioteràpia Global Mézières



Inici de la comercialització al 2014

Tecnologia adoptada per més de 50 clients

50% ROI assolit

Pla de reinversió per al 2017

GRÀCIES

MIGUEL REYES ESTANY

DEPARTAMENT DE MATEMÀTICA APLICADA I ANÀLISI

UNIVERSITAT DE BARCELONA

27 DE FEBRER DE 2017

SUPERVISOR: DR. SERGIO ESCALERA

