

DOCTORADO EN CS. DE LA
COMPUTACIÓN
FACULTAD DE CIENCIAS EXACTAS
UNIVERSIDAD NACIONAL DEL CENTRO DE LA
PROVINCIA DE BUENOS AIRES
TANDIL

TESIS DE DOCTORADO

**Visión computacional y aprendizaje de
máquinas aplicado a la estimación de
activación muscular del bíceps braquial**

Autor:
Leandro M. Abraham ^{1,2,3,4}

Director:
Dr. Facundo BROMBERG^{2,4}
Codirector:
Dr. Raymundo FORRADELLAS³

1-leandro.abraham@frm.utn.edu.ar / leandromaf@gmail.com

2-Laboratorio DHARMa de Inteligencia Artificial,
Departamento de Sistemas de Información,
Facultad Regional Mendoza,
Universidad Tecnológica Nacional

3-CEAL,
Facultad de Ingeniería,
Universidad Nacional de Cuyo

4-Consejo Nacional de Investigaciones Científicas y Técnicas

Resumen

Este trabajo de tesis complementa los requisitos de la carrera de Doctorado en Ciencias de la Computación de la Facultad de Ciencias Exactas de la Universidad Nacional del Centro de la Provincia de Buenos Aires. El problema general en el que se encuentra enmarcada la presente tesis es el de estimar el nivel de activación muscular ejecutado por el bíceps de un brazo humano al ejercer diferentes niveles de esfuerzos discretos, por medio de un sistema de visión artificial alimentado exclusivamente con imágenes externas del brazo. El sistema mencionado tiene embebido un componente de aprendizaje de máquinas supervisado que le permite predecir la medida deseada, a partir de ejemplos conocidos de la activación real. En un primer paso del método, se aplican técnicas de visión computacional para la descripción de imágenes de manera de generar vectores de características de las mismas. Luego, estos vectores son usados como entrada en un proceso de aprendizaje de máquinas supervisado por la fuerza externa aplicada, un valor altamente correlacionado al nivel de activación real que el bíceps está ejerciendo. Se evaluaron técnicas maduras y comúnmente usadas en estas áreas para las cuales existen implementaciones computacionales. Como respuesta a la pregunta de *¿por qué es importante resolver el problema planteado?*, se presenta una introducción al mismo y a las áreas disciplinares de aplicación. Por otro lado, pretendiendo responder *¿qué soluciones existen para el problema planteado?*, se presenta una revisión bibliográfica de los trabajos relacionados al problema y el método propuesto para resolverlo. Para responder a la pregunta de *¿qué debo aprender para resolver el problema mediante el método propuesto?*, se presentan las tecnologías de soporte evaluadas en esta tesis para resolver el problema usando el método propuesto, puntualmente en las áreas de conocimiento de descripción de imágenes y aprendizaje de máquinas supervisado. Finalmente para responder a la pregunta de *¿cómo aplicar estas tecnologías para resolver el problema?*, se presentan distintas instancias del método que pretenden resolver el problema de estimación planteado. Una de las instancias del método presentadas, logra obtener una precisión aceptable en la tarea de medición de la activación en un setting lo suficientemente práctico como para ser usado en ciertos ejercicios de tele-rehabilitación. La precisión en la medición de la activación así como la correlación estimada entre las mediciones obtenidas con este método y la real, son comparables a los métodos actuales y a trabajos similares.

Agradecimientos

A mi esposa Jessica y nuestro hijo Santino por ser soporte incondicional y fuente de motivación constante

A mis padres Jorge y Graciela por su soporte emocional y económico

A mis directores Facundo y Kike por su sabio consejo y guía

A mis hermanos y amigos Magali, Nicolás, Rodrigo y Andrés por su interés y apoyo constante

Contenido

1	Introducción	4
1.1	Introducción	4
1.2	Problema a resolver	4
1.3	El alcance	5
1.4	El enfoque	6
1.5	Estructura del trabajo	7
2	Áreas de aplicación	8
2.1	Introducción	8
2.2	Modelos biomecánicos	8
2.3	Tele-rehabilitación	11
2.4	Electromiografía	11
2.5	Conclusiones	13
3	Trabajos relacionados	14
3.1	Introducción	14
3.2	Trabajos relevados	14
3.3	Conclusiones	16
4	Tecnologías de soporte	17
4.1	Introducción	17
4.2	Aprendizaje de máquinas supervisado	17
4.2.1	Máquinas de vectores de soporte	17
4.2.2	Predictores de bosques aleatorios	19
4.2.3	Medidas de performance	20
4.2.4	Curvas de aprendizaje	21
4.3	Descriptores de imágenes	23
4.3.1	Descriptores de imágenes estáticas 2D	23
4.3.2	Descripción de nubes de puntos 3D	26
4.4	Conclusiones	30
5	Estimación de activación muscular generalizando a nuevos sujetos	31
5.1	Introducción	31
5.2	El enfoque	31
5.3	Evaluación empírica	33
5.3.1	Setup experimental	33
5.3.2	Resultados experimentales	34
5.4	Discusión y conclusiones	38
6	Estimación de activación muscular a partir de modelos especializados para cada sujeto	40
6.1	Introducción	40
6.2	Evaluaciones preliminares usando imágenes RGB como entrada	40
6.2.1	El problema	40
6.2.2	El enfoque	40

6.2.3	Evaluación empírica	41
6.2.4	Discusión y conclusiones	42
6.3	Hacia una aplicación práctica en tele-medicina usando nubes de puntos 3D como entrada . .	43
6.3.1	El problema	43
6.3.2	El enfoque	43
6.3.3	Evaluación empírica	44
6.3.4	Discusión y conclusiones	48
6.4	Conclusiones	53
7	Conclusiones y trabajo futuro	54

Capítulo 1

Introducción

1.1 Introducción

En este capítulo introducimos el problema que se pretende resolver en esta tesis en la Sección 1.2; definimos el alcance de la misma en la Sección 1.3; presentamos en líneas generales el enfoque diseñado para resolver el problema enmarcado en el alcance estipulado en la Sección 1.4; y finalmente delineamos la estructura y contenidos de la tesis en la Sección 1.5.

1.2 Problema a resolver

El nivel de *activación muscular* es una variable de interés tanto en el área de la biomecánica del sistema musculoesquelético, como en otras áreas principalmente relacionadas a la medicina. La solución actual para medir la activación muscular (la electromiografía o EMG [34], presentada en la Sección 2.4) es muy precisa pero requiere equipamiento intrusivo y dependiente de procedimientos de instalación complejos que necesitan personal especializado que adhiera electrodos al cuerpo, limitando la practicidad de su uso. Esta tesis intenta aportar al conocimiento científico y tecnológico con un método que permita la estimación del nivel discreto de activación muscular ejercido por el bíceps de un brazo humano de forma remota, menos intrusiva, mas práctica y económica que las tecnologías actualmente utilizadas para obtener estos datos fisiológicos, restringido a situaciones de contracción estática del músculo sin movimientos que lo estiren o encojan, con la persona relajada para que la contracción muscular sea la mínima requerida para mantener el esfuerzo deseado (conocidas como *contracción isométrica*).

Como se verá en detalle en la Sección 2.2, la biomecánica del sistema musculoesquelético es una disciplina científica que tiene por objetivo el estudio de las estructuras mecánicas, modelos, fenómenos y leyes que son relevantes al equilibrio y movimiento de los seres vivos, principalmente los humanos. Las variables biomecánicas más estudiadas al momento de analizar el equilibrio y movimiento del sistema musculoesquelético humano son las *fuerzas internas* ejercidas por los músculos y los *momentos de fuerza* (en inglés *torques*) que actúan en las articulaciones involucradas. El análisis de estas variables permite identificar movimientos perjudiciales para la salud, situaciones de sobre esfuerzo, posturas (in)adecuadas, trastornos del sistema musculoesquelético, movimientos óptimos, entre otras situaciones del cuerpo humano con alto impacto en su salud y rendimiento. Esto hace que sea muy aplicada en disciplinas como medicina ocupacional [23], ergonomía [57], deportes [64] y tele-medicina, entre otras. En la práctica, la estimación de las fuerzas musculares internas y los torques articulares, se realiza de forma indirecta a través de modelos de dinámica inversa [62]. Las variables de entrada a estos modelos son las posiciones y ángulos articulares junto con las fuerzas externas aplicadas al cuerpo durante el movimiento o postura en estudio. Para capturar las posiciones y ángulos articulares, los dispositivos comúnmente usados son: goniómetros para medir los ángulos (requieren fijar sensores en el cuerpo) y sistemas de captura de movimiento para medir las posiciones articulares visualmente (requieren la adherencia de marcadores y el uso de sistemas multi-cámara caros), entre otros. A pesar de la in-practicidad de los dispositivos clásicos, recientemente la medición de posiciones y ángulos articulares usando cámaras económicas ha alcanzado un nivel importante de madurez gracias a la aparición de sensores de profundidad de bajo costo (por ejemplo Microsoft Kinect TM). Con estos dispositivos, es posible medir las posiciones articulares con una precisión aceptable [37]. Con respecto a la captura de las fuerzas

externas, los dispositivos más comúnmente usados son: plataformas de fuerza para medir las fuerzas de reacción del piso aplicadas sobre los pies y sensores de fuerza para medir las fuerzas aplicadas sobre las manos (los cuales deben ser adheridos a los objetos o en vestimenta instrumentada), entre otros. A pesar de que la medición remota, no invasiva y económica de fuerzas externas mediante imágenes no está aún tan madura, recientemente se han publicado trabajos que intentan estimar las fuerzas de reacción del piso [19], las fuerzas de agarre en las manos [77, 76] y las fuerzas de agarre en las yemas de los dedos [25, 26, 47, 46] a partir de información cinemática y de coloración de la piel capturada con cámaras.

Con respecto a la necesidad de medir la activación muscular en el área de la biomecánica, la misma es útil para superar una de las mayores limitaciones de los modelos biomecánicos. Como se verá en detalle en la Sección 2.2, esta limitación es que los mismos son problemas *ill-posed*, con más de una solución posible donde en la práctica es necesario elegir una de todas estas posibles soluciones y existe la posibilidad de cometer errores en esta elección. Una aproximación común para elegir las fuerzas y torques internos es suponer lo que el cerebro haría, elegir las fuerzas y torques que minimizarían el consumo de energía. Este enfoque sin embargo funcionaría solo cuando el cerebro ha sido entrenado de forma apropiada para el movimiento en estudio y el cuerpo puede responder a lo que el cerebro requiere. En muchas situaciones esto puede no suceder de esta forma ideal, por ejemplo en deportes de alto desempeño, tareas laborales muy demandantes, movimientos cansadores entre otras. La literatura ha propuesto otra solución para este tipo de problemas *ill-posed*: restringir las posibles fuerzas usando información de activación muscular medida directamente en el músculo. Aquí es donde esta tesis pretende hacer una diferencia, proveyendo un método menos invasivo que los actuales para estimar la activación muscular; abriendo la puerta a un sistema en tiempo real de estimación de fuerzas internas y torques en entornos más realistas que los de un laboratorio.

Otra posible aplicación más inmediata del método presentado para medir la activación muscular y que ha sido atacada en esta tesis es en tele-medicina [98, 86, 20, 81] donde se pretende satisfacer la necesidad de medición remota de activación [20, 81, 7]. El EMG como herramienta para medir la activación muscular es muy preciso pero limita la practicidad en ejercicios de tele-fisioterapia. Por el contrario, el método presentado en esta tesis usa equipamiento no intrusivo, tan simple de instalar como una consola de video-juegos, a cambio de menor resolución en la estimación respecto al EMG (ya que solo permite medir niveles discretos de activación). A pesar de su menor resolución, el método presentado presenta precisión equivalente a un EMG, y su resolución es suficiente para algunos ejercicios que requieren monitorear la activación muscular en valores discretos.

Con los problemas presentados en mente, consideramos importante y con gran impacto para la disciplina de la biomecánica y para la medicina en general, poder estudiar la relación entre las características del brazo externamente observables, con el estado de activación interno de los músculos. Incluso sería más importante el poder encontrar una correlación entre estas dos variables de manera de poder lograr mediciones remotas, no invasivas y económicas que puedan ser usadas como información complementaria a los modelos biomecánicos existentes. Contribuir en este camino puede abrir la puerta a una tecnología de medición y análisis biomecánico en tiempo real que pueda ser usada en un mayor rango de escenarios que el de un entorno de laboratorio. Para contribuir a esto, el objetivo de la presente tesis es tratar de validar la siguiente hipótesis:

Existe una función que puede mapear características visuales y geométricas de la superficie del brazo con el estado interno de activación muscular y esa función puede ser aprendida usando visión computacional y aprendizaje de máquinas supervisado.

1.3 El alcance

Debido a la complejidad del problema, fue necesario simplificarlo y acotarlo para que sea alcanzable dentro del periodo de desarrollo de la carrera de doctorado.

Esta tesis propone como simplificación la medición de la activación de forma indirecta y en niveles discretos basada en el hecho de que algunos ejercicios de rehabilitación muscular requieren contracciones isométricas del músculo [55] (presentados con más detalle en la Sección 2.3) para ejercitarlos sin forzarlos a hacer un gran movimiento. Estas son contracciones estáticas del músculo sin movimientos que lo estiren o

encojan, con la persona relajada para que la contracción muscular sea la mínima requerida para mantener el esfuerzo deseado. En tales casos, una fuerza externa tal como el peso de un objeto sostenido por la mano, se transfiere por completo al músculo. Esto da como resultado una correlación creciente directa y monótona entre el peso y el nivel de activación muscular, como se ha demostrado en [59, 85]. Considerando la existencia de esta función y en esta situación específica, la performance obtenida para la estimación de peso, se mantendrá para la estimación de la activación muscular real. Finalmente, en los mencionados ejercicios de contracción isométrica, es suficiente con conocer el nivel discreto de activación del músculo, principalmente para contrastarlo con el valor de la *contracción máxima voluntaria* (MVC del inglés *maximum voluntary contraction*), por lo que el uso de niveles discretos es válido para estos y otros casos de aplicación. Por otro lado, si se pretendiera usar un EMG para medir la activación de referencia, los electrodos taparían la imagen, produciendo ruido adicional que dificultaría el proceso de mapeo. Además, en una aplicación práctica como la presentada anteriormente, este enfoque requiere producir las mediciones para entrenar el algoritmo en escenarios remotos donde el uso de un EMG aumentaría la experiencia requerida por el asistente y/o el paciente.

1.4 El enfoque

Con el fin de aportar en esta área disciplinar y sentar los primeros pasos en el desarrollo de una tecnología que permita en un futuro la medición y análisis biomecánico de forma remota, no intrusiva y económica; es que proponemos un método basado en *clasificación de imágenes* para resolver este problema, como se lo presenta esquemáticamente en la Figura 1.1.

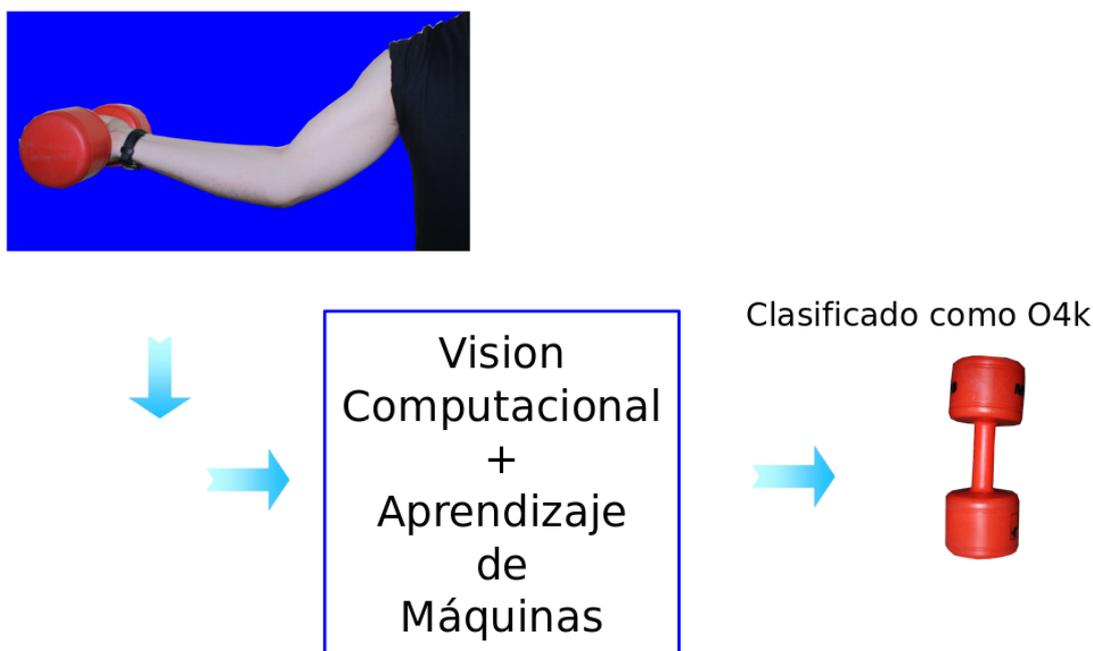


Figura 1.1: El problema se resuelve como uno de clasificación de imágenes

Este método consta de dos etapas bien marcadas:

Calibración o entrenamiento

En esta etapa un gran conjunto de imágenes de cada nivel discreto de activación previamente pre-procesadas son introducidas a algoritmos de visión computacional, los cuales generan un vector de características por cada una de ellas y en su conjunto, agregando la etiqueta del nivel de activación correspondiente, construyen un dataset. Con este dataset se entrenan modelos de aprendizaje de máquinas supervisado para mapear del vector de características al nivel de activación ejercido.

Medición

Esta etapa ocurre cuando es necesario saber cuanto esfuerzo o activación está ejerciendo el brazo del cual se tiene una imagen capturada en las condiciones adecuadas. Esta imagen es pre-procesada, y de ella se obtiene su vector de características para luego evaluar el mismo con el modelo aprendido el cual deberá devolver el nivel de activación ejercido.

Con respecto a las distintas instancias del método experimentadas para resolver este problema, en primer lugar en el Capítulo 5 se experimentó en la resolución del problema pretendiendo generalizar la predicción a sujetos no antes vistos en el entrenamiento. Para este caso, los modelos se entrenaron con información extraída de imágenes RGB de 90 sujetos con características demográficas variadas, y se evaluaron sobre otros 10. Los resultados obtenidos en esta versión de la solución no fueron del todo aceptables debido a la gran complejidad del problema ya que se cuenta solamente con información 2D, y además se cuenta con poca información previa del sujeto lo que hace casi imposible la estimación debido a las características propias de cada persona que no pueden ser generalizables, al menos usando datasets generados con hasta 90 sujetos. En este capítulo se propuso normalizar los datasets usando información del sujeto a evaluar, y esto mejoró los resultados sobre aquellos casos donde no se incluía esta información específica del sujeto. Esto nos permitió entender como la información específica del sujeto a evaluar es muy útil y necesaria y por ello se propuso atacar una versión del problema, más simple pero con impacto aceptable para ciertas aplicaciones médicas. Se pretendió entonces entrenar los modelos para un sujeto en particular como se experimentó en el Capítulo 6. En primer lugar, en la Sección 6.2 se realizó una evaluación preliminar de este escenario, usando nuevamente información 2D RGB. A pesar de los buenos resultados, esperables por la simplicidad del setting de experimentación esa instancia del método carece de aplicabilidad práctica. Para compensar eso, en la Sección 6.3 del mismo capítulo se presenta una instanciación del método que puede generalizar a pequeñas desviaciones en el punto de vista, lo que se logró usando como entrada nubes de puntos 3D y aumentando el dataset con nubes de puntos correspondientes a puntos de vista vecinos espacialmente. Esta simplificación del problema y mejora de la solución logró alcanzar resultados aceptables, y es otro aporte importante que se presenta en esta tesis.

1.5 Estructura del trabajo

En resumen, el resto del documento se organiza del siguiente modo:

- En el capítulo 2 presentamos los conocimientos de fondo de las áreas de aplicación de la biomecánica y la tele-medicina necesarios para entender el problema que pretende resolver esta tesis
- En el capítulo 3 se presentan los trabajos relacionados con el problema a resolver
- En el capítulo 4 presentamos brevemente las tecnologías de visión computacional y aprendizaje de máquinas supervisadas evaluadas en las distintas instancias del método propuesto para resolver el problema planteado
- En el capítulo 5 se evalúa el desempeño de una primer variante de esta metodología para resolver el problema de estimación de activación muscular, generalizando a sujetos no antes vistos durante el entrenamiento de los modelos. Esta variante evalúa el uso de *Patrones binarios locales* e *Histogramas de color HS* como descriptores de imágenes RGB y *bosques aleatorios* y *máquinas de vectores de soporte* como algoritmos de aprendizaje de máquinas
- En el capítulo 6 se evalúa el desempeño de otra variante de esta metodología que propone aprender un modelo específico para un sujeto.
 - En la Sección 6.2 se evalúa el uso de imágenes RGB como entrada, caracterizadas por los descriptores mencionados anteriormente además de la evaluación de *bolsas de puntos claves* y *momentos de contorno*, en conjunto con *máquinas de vectores de soporte* para aprender los modelos.
 - Además en la Sección 6.3 se logra la generalización del modelo a puntos de vista novedosos, usando descriptores globales de nubes de puntos 3D densas (*histograma de features de puntos de vista*, *histograma de features de puntos de vista agrupados* y *conjunto de funciones de forma*) y *máquinas de vectores de soporte* como algoritmo de aprendizaje de máquinas supervisado.
- Finalmente en el capítulo 7 se presentan las conclusiones finales de esta tesis y posibilidades de trabajo futuro.

Capítulo 2

Áreas de aplicación

2.1 Introducción

En este capítulo presentaremos conocimientos de fondo básicos en las áreas de aplicación de la tesis presentada: biomecánica del sistema musculoesquelético en la Sección 2.2, tele-rehabilitación en la Sección 2.3 y electromiografía en la Sección 2.4. Se pretende que estos conocimientos ayuden a comprender mejor el problema que se pretende resolver y ayuden a entender el impacto del problema y de la solución propuesta en esta tesis.

2.2 Modelos biomecánicos

Aquí se explicarán brevemente algunos de los principios considerados para estudiar la biomecánica del equilibrio y movimiento humano que son utilizados para estimar las fuerzas musculares y los momentos de fuerza desarrollados en las articulaciones.

La biomecánica del sistema musculoesquelético es una disciplina científica que tiene por objetivo el estudio de las estructuras mecánicas, modelos, fenómenos y leyes que son relevantes al equilibrio y movimiento de los seres vivos, principalmente los humanos. Las variables biomecánicas más estudiadas al momento de analizar el equilibrio y movimiento del sistema musculoesquelético humano son las *fuerzas internas* ejercidas por los músculos y los *momentos de fuerza* (en inglés *torques*) que actúan en las articulaciones involucradas. El análisis de estas variables permite identificar movimientos perjudiciales para la salud, situaciones de sobre esfuerzo, posturas (in)adecuadas, trastornos del sistema musculoesquelético, movimientos óptimos, entre otras situaciones del cuerpo humano con alto impacto en su salud y rendimiento. Esto hace que sea muy aplicada en disciplinas como medicina ocupacional [23], ergonomía [57], deportes [64] y tele-medicina, entre otras. La estimación de las fuerzas musculares internas y los torques articulares, se realiza a través de modelos de dinámica inversa [62] más procesos de optimización [32, 49, 40] para inferirlas indirectamente. Como su nombre lo indica estos modelos pretenden explicar la dinámica de una postura o movimiento (fuerzas internas y torques) a partir de la cinemática de las posiciones articulares observada. A continuación se explican brevemente estos modelos. Las variables de entrada para estos modelos son: la *cinemática* de las articulaciones en cada instante y las *fuerzas externas* aplicadas al cuerpo durante el movimiento o postura; y la salida de los modelos son las *fuerzas internas* y los *torques articulares*. Una limitación importante de estos modelos de dinámica inversa es que sufren de un problema ill-posed, ya que para estimar las fuerzas musculares que producen los torques articulares, es necesario resolver un sistema con más incógnitas que ecuaciones, debido a que puede haber más de una condición dinámica que haya sido causada por una cinemática observada. Esto se debe al proceso realizado por los mismos:

1. Se empieza tomando como entrada las posiciones y ángulos de los segmentos del cuerpo junto con las fuerzas externas actuando en el movimiento o postura
2. Luego los ángulos articulares se derivan para obtener velocidades y aceleraciones articulares (cinemática) que junto con las fuerzas son introducidas al sistema de ecuaciones de movimiento

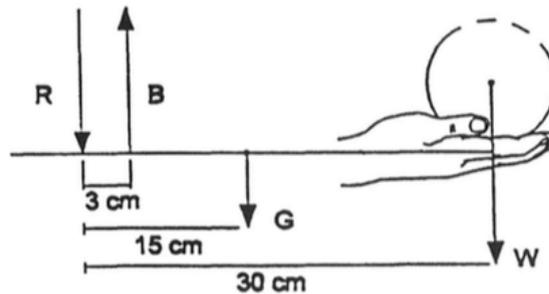
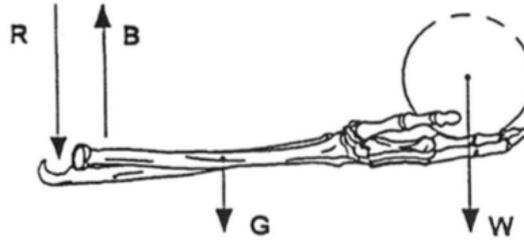
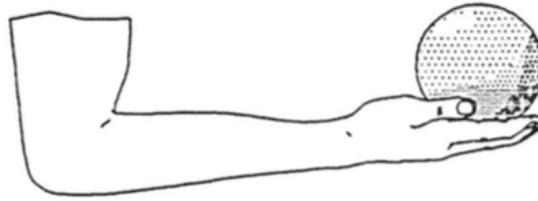


Figura 2.1: Un modelo biomecánico simple de un brazo sosteniendo una bola con el codo flexionado extraído de [57]

3. Finalmente se agrega información inercial y de masa de las partes del cuerpo involucrados para lograr calcular los torques articulares
4. Para estimar las fuerzas musculares internas, las mismas deben ser incluidas en los modelos y se tornan mas complejos por que en el sistema musculoesquelal humano existen múltiples músculos actuando sobre la misma articulación y en consecuencia un gran número de combinaciones de fuerzas musculares que pueden satisfacer las condiciones de equilibrio expresada a través de la cinemática y cinética observada

Para entender mejor los modelos y como se produce el problema ill-posed, presentaremos un modelo biomecánico muy simplificado. Debido a que vamos a trabajar sobre el brazo de un ser humano vamos a presentar un modelo del brazo haciendo un esfuerzo de contracción isométrica al sostener una pelota con la mano y ejerciendo flexión en el codo, extraído de [57]. Este setting es muy similar al que se estudia en esta tesis.

La Figura 2.1 muestra gráficamente este modelo biomecánico estático de un solo músculo del brazo. En este modelo la gravedad actúa sobre la masa de la bola con fuerza de magnitud W , y sobre el centro de masa del antebrazo con una fuerza de magnitud G . Las fuerzas mencionadas actúan a través de sus respectivos brazos de momento de 30 cm y 15 cm de largo desde el punto de flexión del codo. De forma opuesta una fuerza hacia arriba es generada por la contracción del bíceps con una magnitud B que actúa a 3 cm desde

el punto de flexión del codo. Con el fin de mantener el equilibrio, el momento total de flexión/extensión en el punto de flexión del codo debe ser 0 por lo que la condición de momento de equilibrio de la Ecuación 2.1 debe ser satisfecha.

$$(G * 15cm) + (W * 30cm) - (B * 3cm) = 0 \quad (2.1)$$

Como ejemplo, si el antebrazo pesa 10 N y la pelota pesa 15 N, entonces el bíceps deben generar $B = 200$ N para mantener esta postura estática. También las condiciones de equilibrio requieren que la suma de todas las fuerzas en dirección vertical sea cero, satisfaciendo la Ecuación 2.2 donde R es la fuerza de reacción del cuerpo sobre el codo.

$$B - G - W - R = 0 \quad (2.2)$$

En una situación más realista, pueden haber más de un solo músculo actuando sobre una articulación. Suponiendo que otro músculo es agregado a la Figura 2.1 a una distancia de 2 cm desde la articulación y este produce una fuerza H actuando paralelamente al bíceps. En este caso, el sistema de ecuaciones que debe ser resuelto es como el de la Ecuación 2.3.

$$\begin{cases} (G \times 15cm) + (W \times 30cm) - (B \times 3cm) - (H \times 2cm) = 0 \\ H + B - G - W - R = 0 \end{cases} \quad (2.3)$$

En este sistema ahora tenemos tres incógnitas (B , H y R) pero solo dos ecuaciones, lo que resulta en un problema ill-posed. Esto significa que hay un infinito numero de combinaciones de fuerzas musculares que pueden satisfacer las condiciones de equilibrio para esta postura expresada a través de la cinemática y cinética observada.

El enfoque mas usado para mitigar el problema ill-posed y estimar las fuerzas musculares internas es realizar un proceso de optimización estática [32, 49, 40] con el objetivo de encontrar las fuerzas internas musculares mínimas necesarias para satisfacer el equilibrio, siguiendo la lógica de que el sistema nervioso central va a pretender minimizar la energía consumida para ejercer las fuerzas. Para este proceso es necesario proveer además información sobre las propiedades físicas y geométricas del sistema musculoesquelal (largo de los segmentos del cuerpo, largo de los músculos, velocidad de acortamiento de los músculos, volumen muscular), información que en general es extraída de modelos 3D del cuerpo, escalados al sujeto en análisis. De esta forma, las ecuaciones de movimiento deben ser resueltas, sujetas a otras restricciones relacionadas a estas propiedades internas de los músculos y del cuerpo y minimizando una función objetivo que ha sido definida en la mayoría de los trabajos de la literatura considerando las fuerzas internas musculares. Considerando la optimización estática como la técnica más usada para resolver el problema ill-posed de los modelos de dinámica inversa, podemos visualizar una limitación importante. Como se explicó, el problema de optimización va a intentar encontrar las fuerzas mínimas que satisfagan las restricciones, sin embargo esta suposición no siempre es cierta [35] y los humanos tienden a ejercer más fuerza que la mínima estrictamente necesaria para alcanzar el equilibrio o realizar un cierto movimiento. Esto puede suceder por ejemplo cuando se mantienen posturas en condiciones inestables [14]; cuando se manipulan objetos frágiles y se pretende asegurar que el objeto no va a caerse [45]; cuando los músculos están fatigados [39, 73]; por que el control motor esta afectado por la edad [60]; o en la ejecución de movimientos nuevos y desconocidos como en el caso de deportes de alto desempeño o tareas laborales complejas; entre otras posibles causas. En estas situaciones es posible que la solución mínima posible elegida por la optimización sea muy distinta de la real. El nivel de activación muscular juega un papel central para mitigar este problema. Como alternativa, la literatura ha propuesto algunos enfoques híbridos que toman como referencia un modelo basado en activación muscular y optimizan los torques para ser consistentes con las fuerzas musculares estimadas por este modelo [28, 29] o proponen restringir los muchos resultados del problema de dinámica inversa usando información de las fuerzas musculares estimadas por modelos basados en activación muscular [97, 5]. Como se verá más adelante en detalle en la Sección 2.4, hoy en día la medición de activación muscular se realiza con electromiógrafos (EMG), que son dispositivos invasivos que limitan estas mediciones a entornos altamente controlados. Es aquí donde la línea de investigación en la que se enmarca esta tesis hace su contribución distintiva, proporcionando un método no invasivo para estimar esta activación muscular (necesaria para estos modelos); abriendo la puerta a la estimación en tiempo real de las fuerzas y torques en la vida real, fuera de las condiciones del laboratorio, como en entornos de tele-rehabilitación, monitoréo de deportes en vivo o tareas de trabajo in situ.

2.3 Tele-rehabilitación

Además de su importancia en biomecánica, la activación muscular es útil en áreas como la tele-medicina. Puntualmente hemos identificado hasta el momento dos casos de aplicación puntual donde la activación muscular es necesaria en estos entornos y donde la solución propuesta en esta tesis puede ser un aporte de impacto relativamente práctico y rápido.

En primer lugar, los ejercicios *mantener-relajar*, *contraer-relajar* y *mantener-relajar con contracción agonista*, los cuales son comunes en los programas de rehabilitación [55]. En ellos, se requiere monitoreo para controlar la intensidad y duración de la activación muscular, evitando niveles mayores que un porcentaje de la contracción máxima voluntaria (MVC del inglés maximum voluntary contraction, el valor de contracción máximo posible que un sujeto puede producir por si mismo en el músculo) prescrito por el médico. Este monitoreo está acotado a una medición humana e imprecisa de la activación muscular, porque en la mayoría de los casos es imposible operar un EMG para tal escenario de rehabilitación. Cuando no hay un asistente, se requiere que el paciente se preocupe por monitorear el ejercicio, distrayéndolo de la tarea de rehabilitación. El método propuesto en esta tesis y su variante presentada en la Sección 6.3 es adecuado para atacar estos escenarios, donde el sistema puede ser calibrado para detectar la activación máxima indicada para el tratamiento. Esta calibración se puede realizar en presencia de un asistente entrenado durante una sesión en calibración y dejar que el sistema automáticamente detecte sobre-activaciones en sesiones futuras.

Un segundo caso en consideración es una tarea clínica común en rehabilitación que consiste en la evaluación de la fortaleza muscular utilizando el *test muscular manual* [71, 50] (representado gráficamente en la Figura 2.2). Consiste en un sistema de graduación subjetiva en el que el terapeuta mide groseramente el nivel de contracción muscular al tocarlo y sentirlo durante posturas y movimientos específicos. En tele-salud, el contacto directo es imposible, un EMG no es factible, y la transferencia verbal de la medición subjetiva basada en el tacto de un asistente es inexacta. Este método podría hacer posible la ejecución remota de esta prueba calibrando el sistema con la percepción subjetiva de contracción del terapeuta durante una primera sesión cara a cara. Luego, el sistema podría realizar las mediciones de forma remota desde el hogar del paciente.



Figura 2.2: Imagen representativa de un test muscular manual

2.4 Electromiografía

Como es posible ver a partir de las secciones previas, la electromiografía (EMG) es muy útil en biomecánica y medicina debido a que provee una forma precisa de medir activación muscular. El EMG logra tal precisión por que permite grabar y analizar las señales eléctricas producidas por variaciones fisiológicas en el estado de las membranas de las fibras musculares (señales mio-eléctricas) [56].

La información en tiempo real sobre la activación eléctrica de los músculos está relacionada con las fuerzas musculares internas y los momentos de fuerza de las articulaciones [92, 41]. Tanto es así que puntualmente para contracciones isométricas la relación entre EMG y fuerza muscular ha sido considerada como lineal [61, 16, 66] y levemente no lineal [48, 100]. Cuando el largo de las fibras musculares cambia durante la contracción (contracciones dinámicas) esta relación debe ser modelada más conscientemente tomando en

cuenta su relación con cambios en el largo de la fibra muscular [44] y en la velocidad de variación del largo de las fibras [38].

Para capturar la activación eléctrica del músculo, se usan distintos tipos de electrodos. Los tipos de electrodos más usados son conocidos como de *superficie* (ejemplificados en la Figura 2.3) y *percutáneo* (ejemplificados en la Figura 2.4). Los primeros son adheridos sobre la superficie de la piel fuera del músculo para censar la activación eléctrica en esa área. La principal ventaja de este tipo de electrodos es que son menos invasivos e incómodos que los otros. Las principales desventajas que tienen son que no pueden ser usados para medir activación muscular profunda, son menos precisos y es necesario tener en cuenta una gran cantidad de consideraciones en su posicionamiento para obtener una buena señal ya que son muy sensibles a ruidos e interferencias por ejemplo de las señales de otros músculos. Por otro lado, los electrodos *percutáneos* consisten en introducir agujas y/o cables en los músculos profundos. La principal desventaja de este tipo de electrodos es que son muy invasivos y pueden causar incomodidad y dolor al insertarlos. De forma análoga sus ventajas son su precisión, su mejor resistencia a ruidos y la posibilidad de medir activación muscular eléctrica profunda.



Figura 2.3: Ejemplo de electrodos de superficie



Figura 2.4: Ejemplo de electrodos de aguja o percutáneos

Como se ha dejado claro, en biomecánica y medicina la señal de EMG es muy útil [56, 34] ya que brinda información del estado interno del músculo. Además de su uso como entrada en modelos biomecánicos y en tele-medicina, como es presentado en [56] la señal de EMG es usada para responder las siguientes preguntas:

1. ¿Está el músculo activo?: Generalmente esta pregunta es contestada en el nivel nominal por si o por no y puede ser contestada simplemente observando la señal cruda.
2. ¿Está el músculo más o menos activo?: Esta pregunta tiene una respuesta cualitativa obtenida a partir de la discretización de cantidades, por ejemplo: no-activación, moderada, alta. Puede ser respondida comparando amplitudes de EMG.

3. ¿Cuándo el músculo esta encendido o no?: Esta pregunta esta relacionada al tiempo cuando un músculo se activa y puede ser respondida mirando la amplitud de la señal.
4. ¿Cuán activo esta un músculo?: La respuesta a esta pregunta esta dada por un número relacionado con el nivel de activación en alguna escala y puede ser respondida realizando normalización.
5. ¿Está el músculo fatigado?: Esta pregunta puede ser respondida realizando análisis en frecuencia y amplitud específicamente sobre la señal de EMG.

Consideramos que las primeras cuatro preguntas pueden ser respondidas usando el método propuesto en esta tesis con un nivel aceptable de performance. Además de las aplicaciones presentadas anteriormente existen estudios más profundos que se realizan sobre la señal de EMG con el objetivo de encontrar disfunciones musculares y nerviosas. Estos estudios realizan análisis de señales sobre la respuesta eléctrica del músculo y han sido diseñados para trabajar sobre este tipo de señal. Debido a esto es que este tipo de estudios pueden llegar a ser difícil de reproducir usando solo información visual y por ende el método propuesto en esta tesis.

Información más detallada sobre electromiografía puede ser encontrada en [56, 34, 10].

2.5 Conclusiones

En este capítulo se presentaron los conocimientos básicos sobre las áreas de aplicación del método propuesto en esta tesis. Del contenido presentado queda clara la importancia y el impacto de poder medir la activación muscular de forma menos invasiva, más económica y remota que los métodos actuales.

En el próximo capítulo presentamos los trabajos existentes en la literatura relacionados al problema de medición muscular a partir de imágenes corporales externas para entender la originalidad de los aportes de esta tesis.

Capítulo 3

Trabajos relacionados

3.1 Introducción

Hasta donde tenemos conocimiento, no se han encontrado trabajos en la literatura que resuelvan el problema específico de estimar activación muscular a partir de imágenes externas del brazo. Por otro lado, si es posible encontrar trabajos relacionados al problema general de medición de otras propiedades musculares a partir de imágenes corporales. En este capítulo presentaremos estos trabajos y analizaremos su relación con el problema específico que pretendemos resolver.

3.2 Trabajos relevados

A continuación presentaremos los trabajos encontrados en la literatura que utilizan imágenes externas del cuerpo humano para medir propiedades musculares relacionadas a la activación muscular [87, 94, 12, 80, 68, 84, 63, 8, 58, 79, 31, 103, 88, 22, 42, 43, 101, 67, 53, 9, 51, 74, 90]. Presentaremos estos trabajos desde distintos puntos de vista. En primer lugar los analizaremos teniendo en cuenta la tecnología de captura y el formato de las imágenes utilizado. Posteriormente los categorizaremos a partir de las zonas del cuerpo estudiadas. Luego hablaremos sobre la información del cuerpo que mide cada trabajo. Además se analizará cuales son las técnicas de análisis de imágenes utilizadas. Finalmente discutiremos sobre los problemas que se resuelven con cada enfoque.

Es posible identificar varias tecnologías de captura y formato de imágenes predominantes entre estos trabajos. La mayoría de los trabajos utilizan sistemas de captura de movimiento o sistemas de múltiples cámaras costosos [12, 80, 68, 84, 63, 58, 79, 31, 103, 101, 67, 53, 9, 51, 74, 90] debido a su utilidad para obtener información cinemática agregando la mayoría de ellos marcadores para facilitar el seguimiento de las distintas partes del cuerpo o de la piel. Con la reciente aparición de los sensores de color y profundidad (RGB-D) y su capacidad para realizar seguimiento del esqueleto humano en tiempo real, sin marcadores y con relativamente alta precisión, algunos enfoques han utilizado esta tecnología debido a su menor costo económico y practicidad [12, 80]. Otro grupo de trabajos utiliza videos RGB capturados con cámaras de video convencionales [88, 22, 42, 84, 107, 106] o videos RGB-D [43] para capturar información visual cruda en su proceso de sentido muscular. Finalmente se encontró solo un trabajo que utiliza sistemas de escaneo 3D [9]. Con respecto a nuestro enfoque, pretendemos usar dispositivos económicos por lo que se pretende resolver el problema usando imágenes RGB obtenidas con cámaras fotográficas de tipo reflex y nubes de puntos 3D densas capturadas con sensores RGB-D económicos (Microsoft Kinect™).

Entre las zonas del cuerpo medidas o estudiadas por los trabajos presentados podemos categorizarlos en aquellos que se enfocan en las extremidades inferiores [12, 88, 79, 31, 67, 9], en la zona de los brazos [80, 68, 22, 42, 43, 8, 63, 103, 58, 101, 53, 51], pocos se enfocan en todo el cuerpo [84, 74] (aunque con menor nivel de detalle) y en la cara [90]. Por nuestra parte trabajaremos sobre los brazos, puntualmente sobre la zona del músculo del bíceps.

Los trabajos encontrados miden o capturan cierta información del cuerpo y de los músculos estudiados, generalmente para correlacionarla con propiedades de los mismos y estudiar su comportamiento frente a distintas condiciones de esfuerzo. Algunos de los trabajos presentados capturan entre otras variables, información de movimiento del cuerpo, principalmente posiciones articulares y ángulos de los segmentos del

cuerpo [12, 84, 31, 103, 8, 63, 58, 79, 67, 53, 101]. Por otro lado también son medidas variables relacionadas a la deformación producida en la piel debido a la activación muscular interna ejercida [88, 80, 68, 22, 42, 43, 84]. Mediante nuestro enfoque idealmente pretendemos medir variables relacionadas a la deformación de la piel.

Es importante también hablar sobre las técnicas de análisis de imágenes usadas para realizar las mediciones a partir de la información de entrada capturada por los distintos tipos de sensores. La mayoría de los trabajos que miden información cinemática la obtienen mediante un proceso de seguimiento de las partes del cuerpo [12, 8, 63, 58, 79, 31, 101, 67, 53, 9, 74], realizado tanto sobre marcadores mediante sistemas de múltiples cámaras o a partir de sensores RGB-D, como se comentó anteriormente. Por otro lado, las variables relacionadas a la deformación de la piel son obtenidas aplicando tanto técnicas de análisis de deformación de tejidos suaves [88, 22, 42, 43]; reconstrucción de modelos 3D aplicando enfoques basados en datos [80, 68, 84, 31, 51, 9, 74, 90] usada para obtener modelos externos del fenómeno muscular o flujo óptico para medir la magnitud y dirección del movimiento de los píxeles en las imágenes [88]. En cuanto a esta tesis, debido a que trabajamos con imágenes y nubes de puntos estáticas es que evaluamos tanto descriptores de imágenes RGB (*patrones binarios locales, bolsas de puntos clave, momentos de contorno e histogramas de color*) como descriptores globales de nubes de puntos 3D (*histogramas de features por puntos de vista, histogramas de features por puntos de vista agrupados y ensamble de funciones de Forma*).

Los trabajos presentados anteriormente resuelven distintos tipos de problemas relacionados a la medición de activación muscular. De esta manera el trabajo de [12] pretende mostrar de forma interactiva el nivel aproximado de activación o contracción muscular que se producen en los músculos estudiados al realizar ciertos movimientos. Esta información la presentan en una interfaz gráfica de realidad aumentada y el nivel mostrado lo obtienen de una base de datos generada a partir de la medición de EMG y la información cinemática correspondiente. A diferencia de nuestro enfoque en este trabajo utilizan solamente información cinemática de entrada al proceso. Además no pretenden ni logran la estimación de la activación muscular de forma cuantificada si no que se basan en comportamientos de un solo sujeto para mostrar visualmente los patrones aproximados de activación muscular que se producen con cada movimiento, principalmente con un objetivo solo informativo.

Otro problema que han pretendido resolver los trabajos presentados es el de identificación y clasificación de patrones de marcha entre ejemplos de sujetos sanos y de sujetos con alguna disfunción como en [88]. Aquí usan información de deformación de la piel junto con EMG como entrada a redes neuronales, las cuales realizan el proceso de clasificación. Si bien en este trabajo utiliza un enfoque de visión computacional para cuantificar contracción muscular a partir de información de la piel, tampoco aquí pretenden realizar una medición de esta activación, si no más que nada utilizar esta cuantificación como información complementaria a EMG para identificar patrones de marcha que se corresponden a sujetos con disfunciones musculares.

Además varios de los trabajos presentados se han dedicado a resolver el problema de generar modelos 3D de la deformación de la piel debido a activación muscular [80, 68, 84, 9, 51, 74, 90] generalmente con aplicaciones en computación gráfica para animar personajes virtuales. El problema resuelto por estos trabajos difiere del que pretendemos resolver en cuanto a que ellos tampoco pretenden estimar la activación muscular, si no que solamente buscan construir modelos 3D de la deformación debido a la contracción muscular sin una cuantificación o predicción.

Otros trabajos han usado técnicas de visión computacional para evaluar o medir la deformación y movimiento producida en la piel debido a la contracción muscular principalmente usando información de movimiento [22, 42, 43]. Estos trabajos son la aproximación más cercana a la resolución del problema siguiendo un enfoque similar al propuesto por nosotros, ya que utilizan información visual de la piel para caracterizar mediante técnicas de visión computacional la contracción muscular. Sin embargo, los mismos no evalúan esta caracterización contra valores de activación muscular real, ni pretenden usar estas técnicas para una estimación de la misma. Una adaptación [22, 42] para imágenes estáticas fue implementada y evaluada para ser usada en la generación del vector de características, pero la misma no presento buenos resultados por lo que no ha sido incluida en esta tesis. En el caso de que se trabaje con movimiento en algún futuro, puede considerarse su evaluación.

Finalmente un grupo de estos trabajos pretenden resolver el problema de predicción o estimación de activación muscular eléctrica a partir de información cinemática y cinética, la mayoría de ellos validando con capturas de EMG [87, 94, 67, 8, 63, 58, 79, 101, 53] o estimación a partir de un modelo de dinámica inversa [67]. Sobre estos últimos es importante conocer sobre los métodos usados por los mismos para predecir la activación. De esta forma podemos ver que los métodos utilizados incluyen a las redes neuronales [87, 94, 67, 53, 79, 63], métodos probabilísticos de razonamiento bayesiano [53, 8], métodos de ajuste de curvas

[53], modelos de regresión [101] y finalmente modelos biomecánicos [58]. Si bien estos trabajos si resuelven el mismo problema que pretendemos resolver, ellos lo realizan con información cinemática y de fuerzas externas y no a partir de información visual de la piel exclusivamente como se propone en esta tesis. Para estos trabajos, la información cinemática es altamente informativa, sin embargo, podría ser el caso de que para la misma postura o movimiento, el nivel de contracción puede ser diferente, y usar solo la cinemática para inferir la activación puede no ser suficiente y causar mediciones erróneas. Por esta razón, agregan información de fuerza externa como entrada a sus modelos. Aunque la medición de la fuerza directa es útil, en un entorno de medición realista, el uso de sensores de fuerza es un requisito poco práctico que cumplir. Por el contrario, el método propuesto en esta tesis mide el efecto de las fuerzas externas sobre la activación de forma indirecta y no invasiva al considerar la deformación de la superficie del brazo. Además, dado que se estudian posturas estáticas, la información cinemática no es necesaria en las variaciones de nuestro método evaluadas hasta ahora, ya que es un valor constante. Sin embargo, si se analizara movimiento, definitivamente se debe incluir este tipo de información.

3.3 Conclusiones

A partir del relevamiento del estado del arte presentado en este capítulo, es posible ver que no existen antecedentes que realicen estimación de activación muscular a partir de imágenes externas de la piel como es la propuesta en esta tesis. Lo expuesto en este capítulo permite entonces validar la originalidad de la tesis propuesta.

En el próximo capítulo presentamos los conceptos básicos de las tecnologías de soporte usadas y evaluadas con el método propuesto en esta tesis. Estas tecnologías comprenden a los algoritmos de *aprendizaje de máquinas supervisado*, así como las consideraciones metodológicas tenidas en cuenta en el uso de estos algoritmos. Además se explican las técnicas del campo de *visión computacional* para describir imágenes y nubes de puntos con el objetivo de generar vectores de características de las mismas.

Capítulo 4

Tecnologías de soporte

4.1 Introducción

En este capítulo presentaremos las técnicas de las áreas disciplinares aplicadas en el método propuesto para resolver el problema que ataca esta tesis. Como se explicó, el enfoque consiste en dos tareas. El aprendizaje de un modelo utilizando técnicas de aprendizaje de máquinas supervisado que serán explicadas en la Sección 4.2, para lo cual es necesaria la generación de vectores de características a partir de imágenes utilizando descriptores de imágenes los cuales serán explicados en la Sección 4.3.

4.2 Aprendizaje de máquinas supervisado

En esta sección explicaremos brevemente los métodos evaluados para aprendizaje de máquinas supervisado en la tarea de clasificación de imágenes en esta tesis.

El objetivo del aprendizaje supervisado es estimar una función desconocida $\mathbf{y}^*(\mathbf{x})$ la cual representa un conjunto de entrenamiento $T = \{(\mathbf{x}_n, \mathbf{t}_n)\}_{n=1}^N$ de N ejemplos o *data-points*. Los ejemplos se conforman por un vector de variables de entrada \mathbf{x}_n y sus respectivas salidas $\mathbf{t}_n = \mathbf{y}^*(\mathbf{x}_n)$. El objetivo final del proceso de aprendizaje es encontrar una función de hipótesis \mathbf{y} que aproxime correctamente a la función \mathbf{y}^* . Esto último significa que se debe cumplir que $\mathbf{y}(\mathbf{x}_n) = \mathbf{y}^*(\mathbf{x}_n)$ para tantos ejemplos del conjunto de entrenamiento T como sea posible y también para nuevos ejemplos nunca antes vistos. El error cometido por \mathbf{y} con entradas de ejemplos no antes vistos y con salidas desconocidas es llamado error de generalización. Todos los algoritmos de aprendizaje de máquinas supervisados intentan minimizar el error de generalización.

4.2.1 Máquinas de vectores de soporte

En esta sección explicaremos brevemente los fundamentos de un enfoque usado ampliamente para clasificación conocido como *máquinas de vectores de soporte* (SVM por sus siglas del inglés *support vector machines*), el cual fue el que obtuvo los mejores resultados al ser evaluado en el enfoque presentado en esta tesis. Este algoritmo está explicado en [15] y referimos al lector allí para detalles más técnicos.

Suponemos el caso de clasificación binaria donde tenemos un conjunto de *data-points* que pertenecen a una de dos posibles clases. Como cada *data-point* es un vector de n variables, en el caso de SVM como clasificador lineal, se pretende poder separar los *data-points* con un hiperplano de $n - 1$ dimensiones. Como pueden haber una enorme cantidad de posibles hiperplanos que separan a los puntos según su clase, las máquinas de vectores de soporte atacan el problema usando el concepto de margen. El margen es la menor distancia entre el hiperplano separador y los *data-points* más cercanos de cada clase. Este concepto puede ser visto gráficamente en la Figura 4.1. Tal es así que en este enfoque el hiperplano (frontera de decisión) que se elige es aquel para la cual el margen se maximiza. Como se verá posteriormente los *data-points* más cercanos al margen son llamados *support vectors* y son los únicos que se necesitan recordar para lograr la clasificación de nuevos *data-points*.

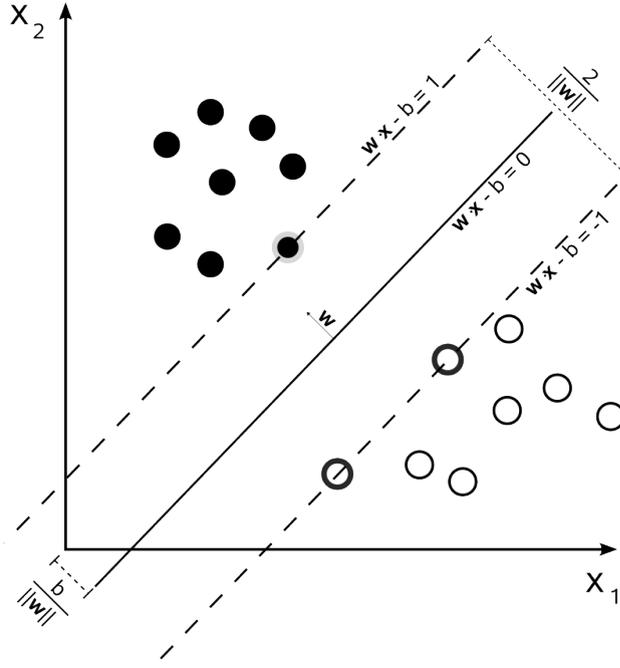


Figura 4.1: Representación del margen

Formalizaremos un poco más puntualmente para el caso de SVM como clasificador lineal. Consideramos que se tiene un conjunto de entrenamiento de n puntos:

$$(\vec{x}_1, y_1), \dots, (\vec{x}_n, y_n)$$

donde y_i puede valer 1 o -1 según a que clase pertenezca el punto \vec{x}_i . Bajo esta convención, cualquier hiperplano puede definirse como el conjunto de puntos \vec{x} que satisfacen

$$\vec{w} \cdot \vec{x} - b = 0$$

Si los datos de entrenamiento son linealmente separables como en la Figura 4.1, es posible elegir dos hiperplanos paralelos que separan los datos de cada clase de tal manera que la distancia entre estos dos hiperplanos (el margen anteriormente mencionado) es la máxima posible. Exactamente a la mitad entre estos dos hiperplanos paralelos se ubica el *hiperplano de margen máximo*. Las ecuaciones de estos dos hiperplanos se ven en la Figura 4.1 y son las siguientes:

$$\vec{w} \cdot \vec{x} - b = 1$$

$$\vec{w} \cdot \vec{x} - b = -1$$

Como se observa en la figura, la distancia entre estos dos hiperplanos está definida por $\frac{2}{\|\vec{w}\|}$ por lo que maximizar el margen se lograría minimizando $\|\vec{w}\|$ sujeto a que ningún data-point caiga dentro del margen y todos ellos estén del lado correcto del mismo. Esto se logra cuando para todo data-point i se cumpla que

$$\text{Si } y_i = 1 \Rightarrow \vec{w} \cdot \vec{x}_i - b \geq 1$$

o

$$\text{Si } y_i = -1 \Rightarrow \vec{w} \cdot \vec{x}_i - b \leq -1$$

o lo que es lo mismo:

$$y_i * (\vec{w} \cdot \vec{x}_i - b) \geq 1 \quad \text{para todo } 1 \leq i \leq n \quad (4.1)$$

Lo que se busca entonces es minimizar $\|\vec{w}\|$ sujeto a las restricciones de la Ecuación 4.1. Este pequeño desarrollo permite ver que una vez que se encuentra el valor de \vec{w} y b adecuados los únicos data-points del conjunto de entrenamiento necesarios para clasificar nuevos data-points son los que están sobre los hiperplanos paralelos separadores que como mencionamos son llamados *support vectors*.

Cuando los datos de entrenamiento no son linealmente separables, es posible usar clasificadores no-lineales, los cuales aplican lo que se conoce como el *truco del kernel* (*kernel trick* en inglés). Este "truco" consiste en aplicar una función de kernel no lineal que transforma los data-points a un espacio de features de mayor dimensionalidad donde los mismos si son linealmente separables.

Habiendo visto los fundamentos teóricos básicos de este enfoque podemos hablar sobre clasificación multi-clase. Para esta tarea, se han propuesto distintos métodos que combinan múltiples clasificadores SVM binarios.

Un enfoque comúnmente usado es conocido como *uno-versus-el-resto* y consiste en construir K SVMs por separado en los cuales el k -ésimo modelo es entrenado usando datos de la clase c_k como ejemplos positivos y el resto como ejemplos negativos. Este enfoque tiene el problema de que puede llevar a resultados inconsistentes donde una entrada es asignada a más de una clase. Esto puede ser resuelto realizando las predicciones usando

$$y(x) = \max_k y_k(x) \quad (4.2)$$

Otro enfoque llamado *uno-versus-uno*, consiste en entrenar $K(K-1)/2$ SVM binarios diferentes sobre todos los pares posibles de clases y luego clasificar los puntos nuevos de acuerdo a cual clase tiene la mayor cantidad de votos.

4.2.2 Predictores de bosques aleatorios

Los *bosques aleatorios* (conocidos en inglés como *random forests*) [17] son una técnica de aprendizaje de máquinas supervisado relativamente reciente que es usada para clasificación con buenos resultados, y la misma también fue evaluada en el contexto de esta tesis en la tarea de discriminación de imágenes RGB en base a niveles de activación muscular del bíceps. Bajo la idea de combinar predictores es que esta técnica aprende una serie de árboles de decisión [18] y a partir de estos árboles aprendidos es que para predecir un nuevo ejemplo se predice con cada uno por separado y luego mediante un proceso de votación se elige el resultado que tiene mayor cantidad de votos. Algunas de las particularidades de este enfoque tienen que ver con: (a) la posibilidad de manejar muchas variables; (b) el poder dar una estimación de la importancia de cada variable en la predicción; y (c) que sirven para detectar interacciones entre variables.

El proceso de entrenamiento sigue los siguientes pasos para cada árbol a entrenar.

1. Muestrear N casos del conjunto de entrenamiento al azar con reemplazo para crear un subconjunto de estos datos
2. En cada nodo:
 - (a) Se seleccionan al azar siguiendo una distribución definida un número m de variables (menor a la cantidad total) para particionar el conjunto
 - (b) La variable predictora del conjunto anterior que provee la mejor división del conjunto de acuerdo a una función objetivo se usa para realizar una división binaria en ese nodo
 - (c) Para el próximo nodo se seleccionan otras m variables predictoras al azar y con la misma distribución y se vuelve a iterar
3. El árbol se hace crecer lo más posible, sin realizar poda

Luego de este proceso de entrenamiento se obtienen una serie de clasificadores (o regresores) entrenados. Cuando se necesita predecir un valor para un nuevo ejemplo, se recorre cada árbol hasta las hojas y se decide por la etiqueta que tenga la mayor cantidad de ocurrencias como se muestra de forma simplificada en la Figura 4.2.

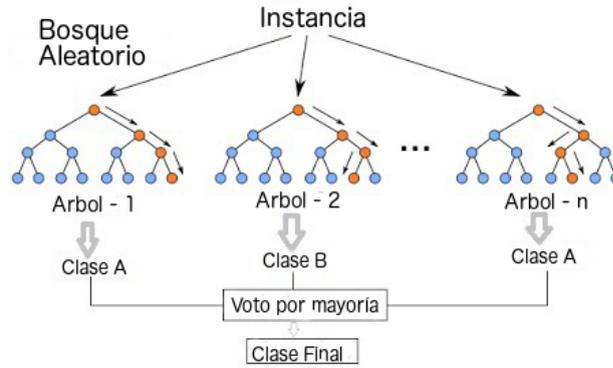


Figura 4.2: Representación simplificada del proceso de clasificación con random forests

Una definición más formal, extraída de [17], dice que:

Un *bosque aleatorio* es un clasificador que consiste en una colección de clasificadores de tipo árbol $\{h(\mathbf{x}, \Theta_k), k = 1, \dots\}$ donde $\{\Theta_k\}$ son vectores elegidos al azar, independientes e idénticamente distribuidos y cada árbol emite un voto unitario para la clase más popular para la entrada \mathbf{x} .

Como información adicional, referimos al lector a [17].

4.2.3 Medidas de performance

En este trabajo, informamos el error de generalización de los modelos entrenados con algoritmos de clasificación mediante el uso de medidas clásicas de rendimiento para esta tarea: *accuracy*, *precision*, *recall* y *f-measure*. Para explicar cada una de ellas debemos recordar primero el concepto de *matriz de confusión* (como la de la Figura 4.3), que en su celda (i, j) muestra el recuento C_{ij} de data-points en el conjunto evaluado que son etiquetados con la etiqueta i (ground truth) pero fueron clasificados por el modelo entrenado como la clase j (prediction). Todas las celdas en la diagonal representan las clasificaciones correctas (es decir, una coincidencia entre la etiqueta verdadera y la que produce el modelo entrenado), mientras que los elementos que no están en la diagonal son errores de clasificación.

		Predicción				
		1	...	j	...	n
Etiqueta Real	1	C_{11}		C_{1j}		C_{1n}
	...					
	i	C_{i1}		C_{ij}		C_{in}
	...					
	n	C_{n1}		C_{nj}		C_{nn}

Figura 4.3: Representación general de una matriz de confusión, donde cada celda (i, j) muestra la cuenta C_{ij} de data-points etiquetados con la clase i , que fueron clasificados como pertenecientes a la clase j

La Accuracy sobre toda la matriz de confusión, nombrada como *Overall Accuracy* y denotada por OvA , se calcula como la suma de los recuentos en la diagonal, dividido por la suma de los recuentos de todas las

celdas. Para casos con número de clases $n > 2$, como es el caso en la presente tesis; la *precision*, *recall*, *f-measure*, y *accuracy* se pueden calcular para cada clase de forma independiente, lo que resulta en n de ellos. Siguiendo la práctica habitual, informamos su media sobre todas las clases. Para cada clase, denotamos cada una de ellas como P_{clase} , R_{clase} , FM_{clase} , y A_{clase} , y sus valores promedio para todas las clases como \hat{P} , \hat{R} , \hat{FM} y \hat{A} , respectivamente; y se calculan de la siguiente manera:

- P_j como el número de ejemplos clasificados correctamente de la clase j dividido por el número total de ejemplos clasificados como clase j
- R_i como el número de objetos clasificados correctamente de la clase i dividido por el total de ejemplos que realmente son de la clase i
- FM_i como combinación de precisión y recall de la clase i
- Accuracy A_i como la suma de las cantidades de la diagonal + la suma de las cantidades que no pertenecen ni a la fila i ni a la columna i , dividida por el número total de ejemplos.

Además de estas medidas de rendimiento, decidimos informar también el *error absoluto medio* (MAE del inglés *mean absolut error*) que contempla el valor en gramos de la etiqueta de clase, y se calcula como la media de los valores absolutos de las diferencias entre la fuerza externa aplicada correspondiente al nivel de activación estimado, menos la fuerza externa aplicada correspondiente al verdadero nivel de activación. A partir de esta definición, mejores rendimientos se reflejan en valores de MAE más pequeños. El beneficio de esta medida es que pondera los errores de manera diferente según su distancia a su verdadero valor. En cambio, en todas las medidas de desempeño consideradas anteriormente, los errores se ponderan simplemente como 1 en cada conteo del matriz de confusión. Las fórmulas de todas estas medidas se presentan con más detalle en la Tabla 4.1.

Medida	Formula
OVA	$\frac{\sum_{i=1..n} C_{ii}}{\sum_{k=1..n} \sum_{l=1..n} C_{kl}}$
P_j	$\frac{C_{jj}}{\sum_{i=1..n} C_{ij}}$
\hat{P}	$\frac{\sum_{c \in \text{Clases}} P_c}{ \text{Clases} }$
R_i	$\frac{C_{ii}}{\sum_{j=1..n} C_{ij}}$
\hat{R}	$\frac{\sum_{c \in \text{Clases}} R_c}{ \text{Clases} }$
FM_i	$2 \times \left(\frac{P_i \times R_i}{P_i + R_i} \right)$
\hat{FM}	$\frac{\sum_{c \in \text{Clases}} FM_c}{ \text{Clases} }$
A_i	$\frac{C_{ii} + \sum_{l=1..n, l \neq i} \sum_{k=1..n, k \neq i} C_{lk}}{\sum_{i=1..n} \sum_{j=1..n} C_{ij}}$
\hat{A}	$\frac{\sum_{c \in \text{Clases}} A_c}{ \text{Clases} }$
MAE	$\frac{\sum_{d \in \text{data-points}} \text{Predictedg}(d) - \text{Realg}(d) }{ \text{data-points} }$

Tabla 4.1: Medidas de performance evaluadas

4.2.4 Curvas de aprendizaje

Presentamos en esta sección una explicación breve de las *curvas de aprendizaje* [75, 6], una herramienta muy utilizada en proyectos de aprendizaje de máquinas para evaluar en más detalle el desempeño de los modelos en caso de que los mismos no presenten resultados aceptables. Estas curvas consisten en comparar la evolución del error de generalización y el error de clasificación producido por el clasificador cuando se lo entrena con una cantidad creciente de ejemplos. La comparación del error de generalización y el de entrenamiento provee

indicios del potencial de posibles reducciones en el error de generalización de los modelos al entrenarlos con mas ejemplos. La principal intuición es que con conjuntos de entrenamiento pequeños, el algoritmo puede producir modelos lo suficientemente complejos para modelar las particularidades del ruido en los datos, lo que comúnmente se conoce como *over-fitting*. En contraste, para conjuntos de entrenamiento lo suficientemente grandes, el espacio de modelos encuentra sus limitaciones en la complejidad de los modelos que puede proveer, teniendo que negociar un modelo mas simple que las complejidades del ruido de un dataset tan grande. En el primer caso de un training-set pequeño, los resultados del modelo aprendido presentan errores de entrenamientos bajos (ya que se modela exactamente los datos), pero los errores de testeo son muy grandes y alejados del error de entrenamiento. Al aumentar los datos, el error de entrenamiento tiende a aumentar levemente, pero el de generalización empieza a disminuir y ambos tienden a juntarse en un punto medio. Las curvas de aprendizaje ideales generalmente empiezan con un error grande de generalización y un error de entrenamiento muy pequeño (incluso cero), y para mayor cantidad de ejemplos de entrenamiento tienden a converger a un error común.

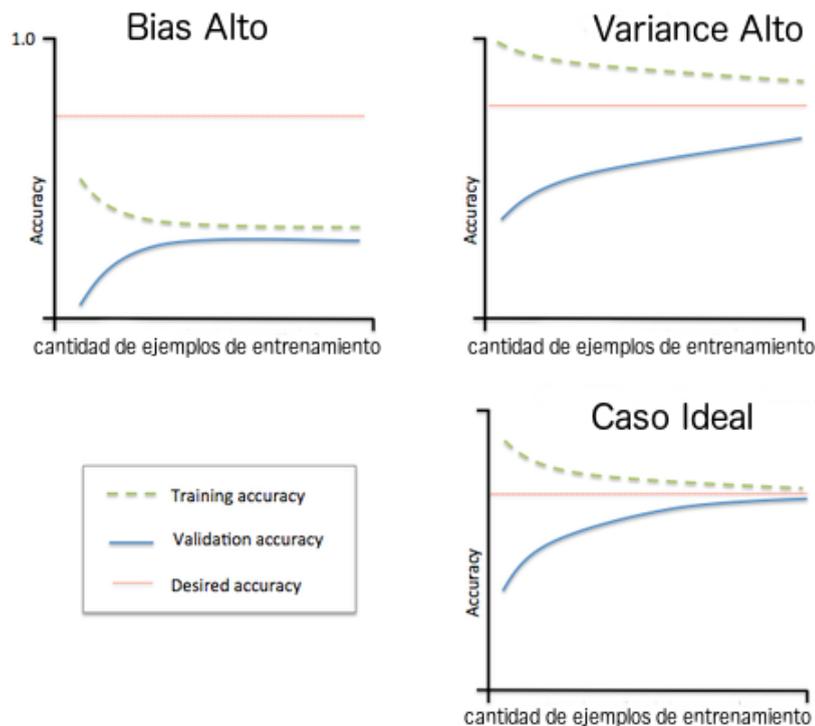


Figura 4.4: Interpretación del error en curvas de aprendizaje

Explicaremos un procedimiento práctico para analizar estas curvas, basándonos en la Figura 4.4. Según el problema a atacar, se define la performance deseada o esperada para ese problema (*Desired accuracy*, línea rellena roja en la figura). Generalmente esta performance esta dada por la performance esperada para un ser humano en la misma tarea. En primer lugar se analiza la performance de entrenamiento (línea cortada verde en la figura) principalmente para la zona de la curva con más ejemplos de entrenamiento y puede suceder lo siguiente:

1. La performance de entrenamiento es mucho peor que la performance deseada, como se observa en la parte superior izquierda de la figura, en cuyo caso estamos ante un problema de bias alto o *under-fitting*. En este caso generalmente la solución es proponer modelos más complejos.
2. La performance de entrenamiento es parecida o igual a la deseada. En este caso, hay que analizar la diferencia de la performance de entrenamiento con la de validación o testeo (línea rellena azul en la figura) y pueden darse dos situaciones:
 - (a) La performance de validación está muy alejada de la de entrenamiento, como se observa en la parte superior derecha de la figura, en cuyo caso estamos en una situación de variance alto u

over-fitting. Aquí la solución generalmente es conseguir más datos de entrenamiento

- (b) La performance de validación es cercana a la de de entrenamiento, como se observa en la parte inferior derecha de la figura, en cuyo caso estamos en el caso ideal

4.3 Descriptores de imágenes

En esta sección vamos a presentar brevemente las técnicas de visión computacional evaluadas en este trabajo para representar imágenes como vectores de características, los cuales son la entrada de los algoritmos de aprendizaje de máquinas.

4.3.1 Descriptores de imágenes estáticas 2D

En los próximos párrafos presentaremos los descriptores de imágenes 2D evaluados en esta tesis describiendo brevemente como funcionan.

Patrones binarios locales (LBP)

Una forma de extraer características de imágenes RGB muy usada principalmente para clasificación de textura son los *patrones binarios locales* (LBP del inglés *local binary patterns*) [69, 70, 105]. Esta técnica también ha sido usada ampliamente para caracterización de la piel en el campo de reconocimiento de expresiones faciales [89, 104]. Debido a su buen desempeño para caracterizar cambios en texturas es que se optó por evaluar este descriptor en esta tesis en el Capítulo 5 y la Sección 6.2 del Capítulo 6.

Dada una región de interés (ROI) en escala de grises, el operador básico genera un código binario para cada píxel de la ROI. Este código se obtiene comparando el píxel central g_c con sus P píxeles vecinos circulares g_p con $p = 0, \dots, P - 1$ en un radio R y decidiendo el valor 0 si $g_c > g_p$ o 1 en el caso contrario. A continuación, el código binario se convierte a su valor decimal correspondiente. El operador básico funciona en $P = 8$ píxeles vecinos en un radio $R = 1$ y esos son los valores evaluados en esta tesis. En [70] el operador $LBP_{P,R}$ es presentado formalmente como en la Ecuación 4.3 y este operador permite obtener los códigos explicados anteriormente.

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) \quad (4.3)$$

donde

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \quad (4.4)$$

Finalmente los códigos generados por el operador para cada píxel de la ROI son cuantificados en un histograma, que es el vector final de la ROI.

Un alternativa simplista y comúnmente usada para encontrar las regiones de interés, la cual se utilizó en esta tesis, es dividir la imagen a la que se desea extraer un vector de características en celdas. Luego se calcula LBP para cada región muestreada. Finalmente se combinan sus vectores de características locales en uno global. Todo este proceso se puede ver gráficamente en la Figura 4.5.

Histogramas de color HS

Para caracterizar una región de una imagen también es posible hacerlo a partir de su información de color. Una forma ampliamente utilizada para describir un conjunto de píxeles es a través de histogramas que cuentan la cantidad de píxeles que caen en distintos rangos del espacio de colores. La opción de histograma evaluada en esta tesis en el Capítulo 5 y la Sección 6.2 del Capítulo 6, fue describir el conjunto de píxeles a través del histograma del espacio de color HSV, que mejora la invarianza con respecto a la intensidad de la luz. Esta técnica ha sido utilizada en tareas similares por ejemplo para clasificar imágenes en conjunto con SVM [24] y para caracterizar las imágenes en sistemas de recuperación de imágenes [93]. Además se ha utilizado ampliamente para caracterizar piel. Para esta tesis se considero solamente usar los canales H y S para construir los histogramas.

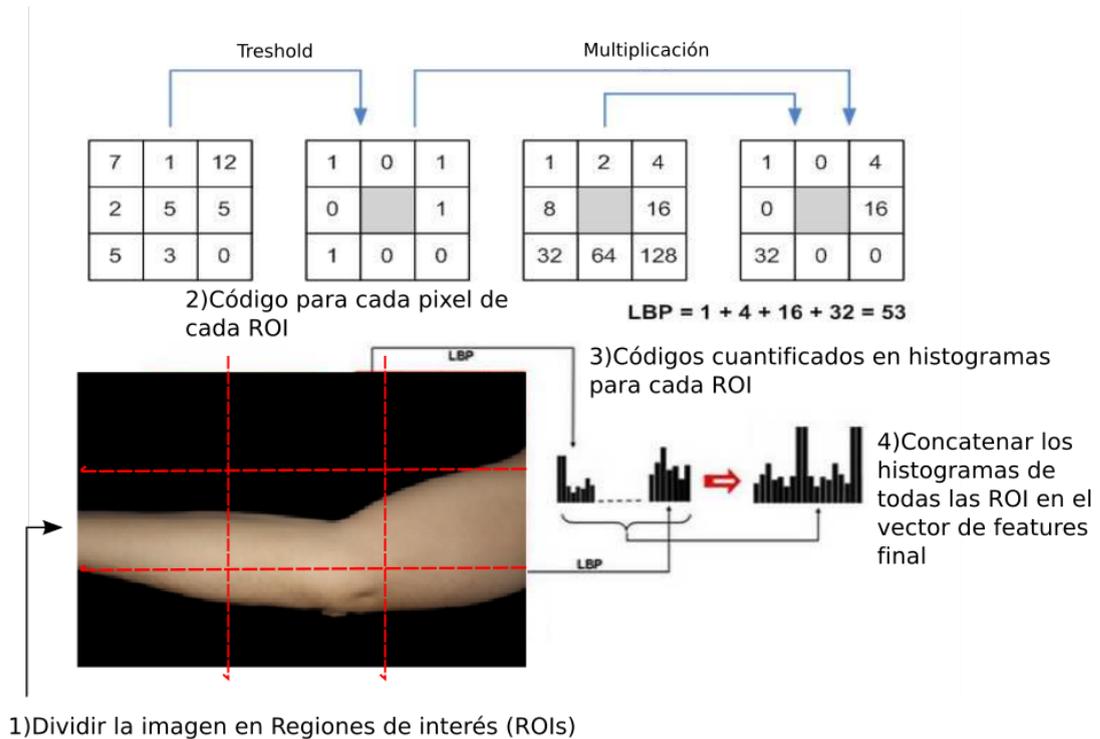


Figura 4.5: Ejemplo del algoritmo *LBP* aplicado

La implementación evaluada de esta técnica en esta tesis en primer lugar divide la imagen a caracterizar en una grilla de regiones de interés. Luego cada grilla es transformada al espacio de colores HSV y cada uno de los píxeles es analizado para evaluar sus valores y sumar al *bin* correspondiente del histograma correspondiente. Luego de obtener un histograma para cada canal y para cada región de interés, los histogramas de cada ROI son concatenados en un vector final global de la imagen. Todo este proceso se puede ver gráficamente en la Figura 4.6.

Bolsa de puntos clave

Otro enfoque para extraer características visuales ampliamente utilizado para resolver el problema de clasificación de imágenes es el de bolsa de puntos clave (BKP del inglés *bag of keypoints*) [33]. Este descriptor fue evaluado en esta tesis en la Sección 6.2 del Capítulo 6.

El enfoque puede ser resumido en un proceso de varios pasos el cual también está representado gráficamente en la Figura 4.7:

1. Entrenamiento:
 - (i) Detección de puntos clave: En este paso se encuentran puntos clave en las imágenes. Los puntos claves o *keypoints* Son porciones de la imagen que presentan características que pueden ser igualmente detectadas bajo variación de escala, iluminación o ruido y generalmente se caracterizan por ser zonas de gran contraste en la imagen. En esta tesis utilizamos el algoritmo SURF [13] para encontrar estos puntos de interés.
 - (ii) Descripción de puntos clave : Los puntos clave encontradas se caracterizan utilizando un vector que contiene información de la posición, la orientación, los gradientes de la imagen en esa zona, entre otros.
 - (iii) y (iv) Entrenar una técnica de agrupamiento para cuantificar los puntos clave de todas las imágenes en un diccionario: Los puntos clave se agrupan según sus características en un número definido de conjuntos (palabra del diccionario). En esta tesis el agrupamiento se obtuvo aplicando la técnica de agrupamiento *KMedias* [15] a los vectores descriptivos de los puntos de interés.

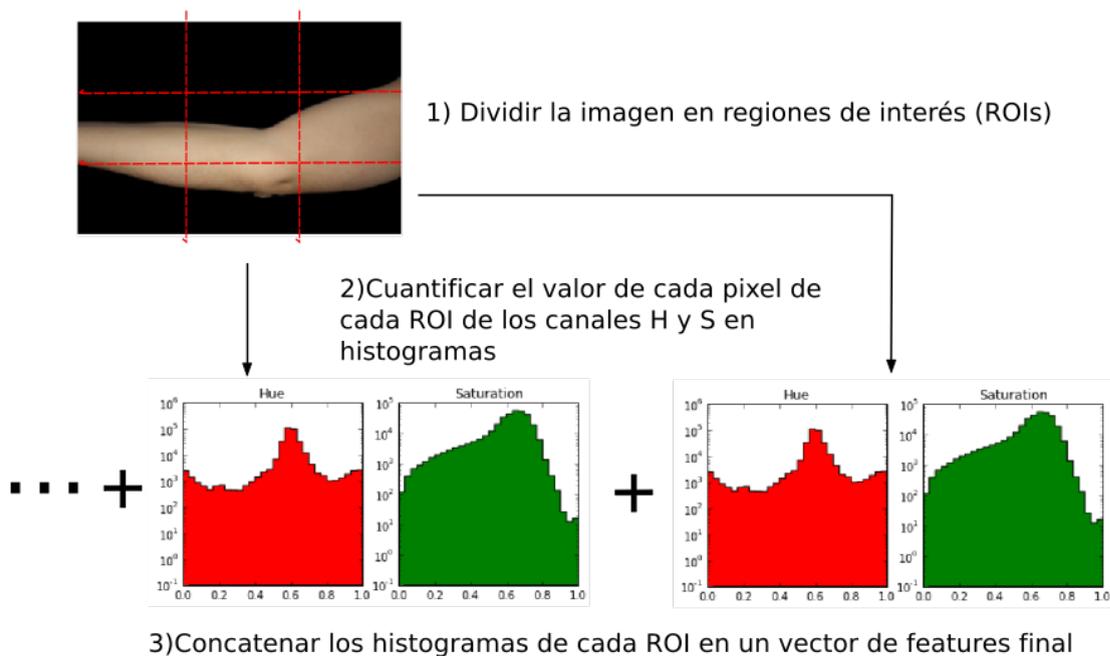


Figura 4.6: Ejemplo del algoritmo *Histogramas de color HS* aplicado

2. Generación de vectores de características:

- (v) Detección de puntos clave en la imagen a evaluar
- (vi) Generación de vectores de características para cada punto clave
- (vii) Asignar cada punto clave a un conjunto para obtener un histograma de ocurrencia de cada palabra del diccionario en cada nueva imagen (bolsa de palabras visuales)
- (viii) Devolver estos histogramas como vectores de características

Momentos de contorno

Otra forma de caracterizar una imagen es usando información de forma a través de su contorno. Una de las formas más simple y mas usada para comparar contornos es a través de sus momentos. El momento de un contorno es el promedio pesado de las intensidades de los píxeles. El momento $m_{p,q}$ de un contorno se define como en la Ecuación 4.5.

$$m_{p,q} = \sum_{i=1}^n I(x,y)x^p y^q \quad (4.5)$$

Donde p y q son conocidos como ordenes del momento. La sumatoria es sobre todos los píxeles del contorno (n) y los contornos de la imagen pueden ser obtenidos usando un algoritmo de detección de bordes.

Es posible definir distintos tipos de momentos. Los momentos espaciales son 10 y pueden ser calculados variando los ordenes de x y y (valores p y q) en la Ecuación 4.5.

Los momentos centrales son 7 y ellos pueden obtenerse sumando información de los momentos de orden 0 y 1 en la fórmula original como es posible ver en la Ecuación 4.6

$$\mu_{p,q} = \sum_{i=0}^n I(x,y)(x - x_{avg})^p (y - y_{avg})^q \quad (4.6)$$

donde $x_{avg} = m_{10}/m_{00}$ y $y_{avg} = m_{01}/m_{00}$

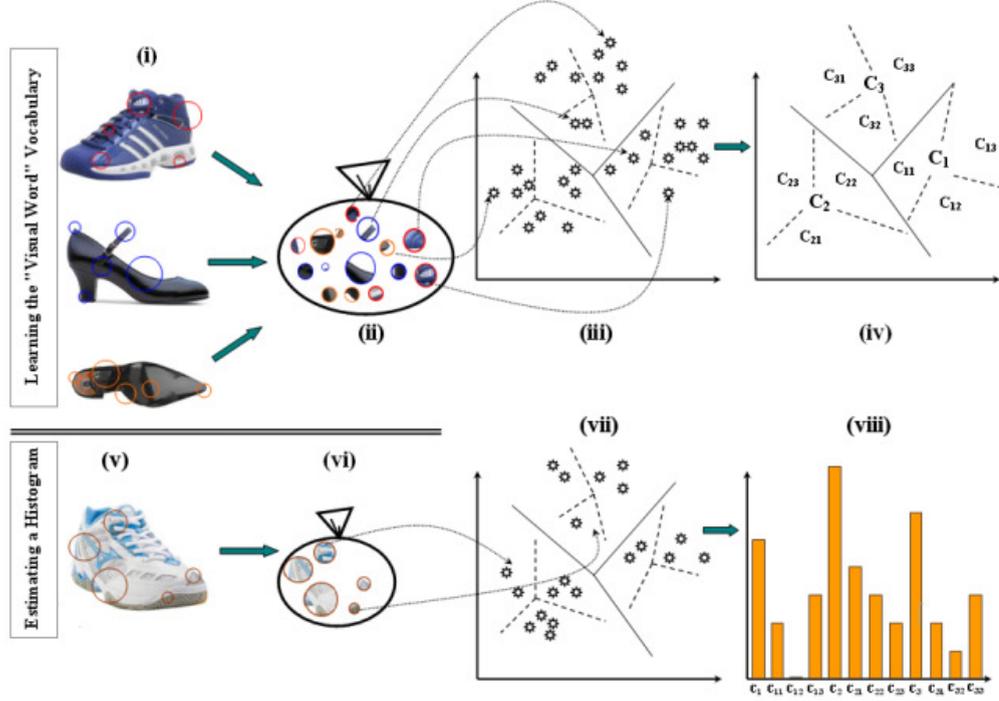


Figura 4.7: El enfoque *Bolsa de puntos clave* resumido

Los momentos centrales normalizados son 7 y pueden ser estimados a partir de los momentos centrales divididos por una potencia del momento m_{00} como se puede observar en la Ecuación 4.7

$$\eta_{p,q} = \frac{\mu_{p,q}}{m_{00}^{(p+q)/2+1}} \quad (4.7)$$

Finalmente los momentos invariantes de Hu [52] son 7 y son calculados como combinaciones lineales de los momentos centrales normalizados como podemos ver en las siguientes ecuaciones

$$\begin{aligned} h_1 &= \eta_{20} + \eta_{02} \\ h_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \\ h_3 &= (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\ h_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} - \eta_{03})^2 \\ h_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})((\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2) + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})(3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2) \\ h_6 &= (\eta_{20} - \eta_{02})((\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2) + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\ h_7 &= (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})(3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2) - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})(3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2) \end{aligned}$$

Estos últimos son muy utilizados debido a que caracterizan los contornos de forma invariante a escala, rotación y reflexión.

Concatenando todos estos momentos es posible generar un vector de $n = 10+7+7+7 = 31$ características, el cual fue evaluado en esta tesis en la Sección 6.2 del Capítulo 6.

4.3.2 Descripción de nubes de puntos 3D

Para caracterizar nubes de puntos 3D existen también en la literatura una gran variedad de técnicas. De estas técnicas los comúnmente conocidos como *descriptores globales* toman como entrada una nube de puntos 3D de un objeto completo y generan el vector de características para todo el objeto. En esta sección presentamos 3 descriptores globales de nubes de puntos utilizados para generar un vector de características a partir de una nube 3D de un brazo, evaluados en esta tesis en la Sección 6.3 del Capítulo 6.

Histograma de features de puntos de vista (VFH) e Histograma de features de puntos de vista agrupados (CVFH)

El descriptor conocido como VFH (por sus siglas en inglés *viewpoint feature histogram*) ha sido descrito en detalle en [82]. El mismo representa las orientaciones relativas de las normales y las distancias entre cada punto de la nube y el centroide de la misma. Cada uno de estos ángulos y distancias son contabilizados en un histograma.

Este descriptor caracteriza una nube de puntos con información sobre la orientación relativa de la normal y de la distancia entre cada punto p_i con el centroide de la nube p_c . En base a estos puntos se construye un sistema de coordenadas que se define utilizando la superficie normal n_i de p_i , el vector $(p_c - p_i)/(\|p_c - p_i\|_2)$ y el producto cruz de estos dos vectores como se muestra en las siguientes ecuaciones:

$$\begin{aligned} u &= n_i \\ v &= u \times \frac{p_c - p_i}{\|p_c - p_i\|_2} \\ w &= u \times v \end{aligned}$$

Basado en el sistema de coordenadas que forman estos vectores, los ángulos de orientación de la normal α, ϕ, θ se estiman usando las siguientes ecuaciones:

$$\begin{aligned} \alpha &= v^T \cdot n_i \\ \phi &= u^T \cdot \frac{p_c - p_i}{\|p_c - p_i\|_2} \\ \theta &= \arctan(w^T \cdot n_i, u^T \cdot n_i) \end{aligned}$$

Estos cálculos se ven representados gráficamente en la Figura 4.8

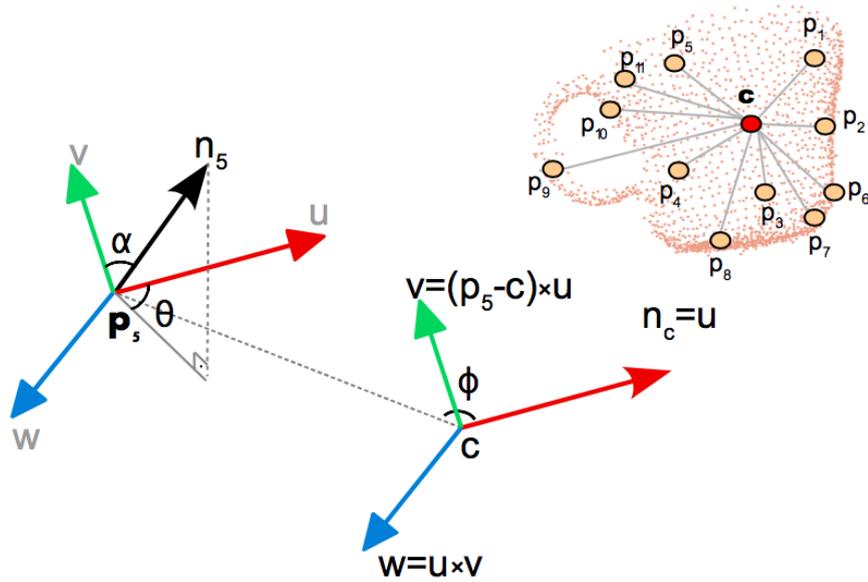


Figura 4.8: Cálculo de ángulos del descriptor VFH extraído de [82]

Además se agrega al vector un componente relativo al punto de vista de la captura que se computa mediante un histograma de los ángulos que el punto de vista forma con cada normal.

Otro descriptor evaluado en esta tesis es el conocido por sus siglas CVFH (del inglés *clustered viewpoint feature histogram*) y que ha sido publicado en [4]. El mismo trabaja dividiendo el objeto en N regiones suaves disjuntas y para cada una de estas regiones se computa un descriptor VFH. El enfoque detrás de CVFH es usar las partes del objeto para construir el sistema de coordenadas, pero usando la vista completa del objeto

para computar el descriptor. Además este descriptor agrega a sus variables un componente de distribución de forma que codifica información de la relación de cada punto con el centroide de su correspondiente región en el objeto como se observa en la siguiente ecuación usada para calcularlo:

$$SDC = \frac{(p_c - p_i)^2}{\max((p_c - p_i)^2)}$$

Conjunto de funciones de forma (ESF)

El descriptor ESF (del inglés *ensemble of shape functions*) [99] también fue evaluado en esta tesis y es el que mejor resultados presenta. El mismo es un conjunto de diez histogramas que suman funciones describiendo propiedades de forma y ángulos característicos de la nube de puntos, las funciones de forma que conforman el descriptor son las siguientes.

Un grupo de variables es estimado calculando la función de forma $D2$ la cual se obtiene muestreando pares de puntos de la nube completa y construyendo un histograma de distancias entre ellos, como se observa gráficamente en la Figura 4.9

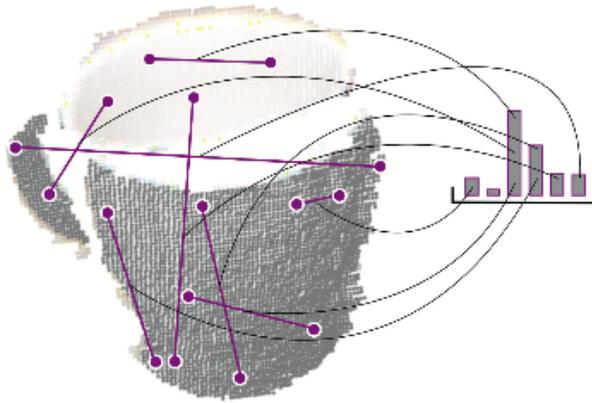


Figura 4.9: Función de forma $D2$ de ESF extraído de [99]

Otro grupo de variables del vector completo son estimadas extendiendo $D2$ trazando líneas entre los pares de puntos muestreados al azar y sumándolos en un histograma que cuenta la cantidad de líneas que caen *en* la superficie de la nube, cuantas *fuera* de ella y cuantas tanto *dentro como afuera* (denominadas como *mixed*). Esta función de forma se ve gráficamente en la Figura 4.10.

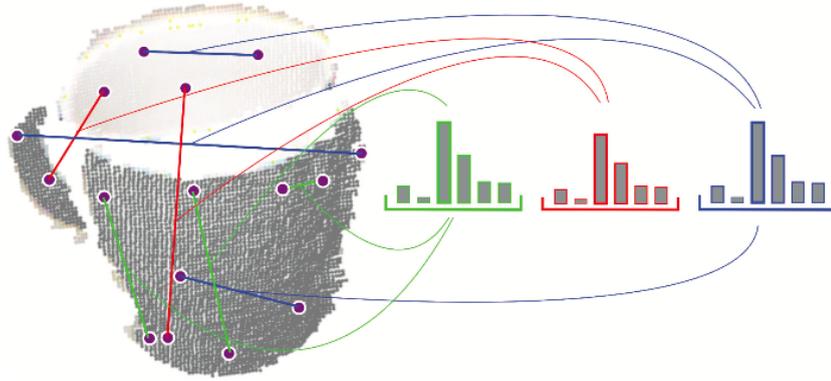


Figura 4.10: Función de forma extensión de $D2$ de ESF extraído de [99] que cuantifica la distribución de líneas sobre la superficie de la nube

Derivado de esta clasificación de las líneas, tomando solo las que se consideran *mixed*, sus ratios son cuantificados en otro histograma que forma el descriptor final, como se ve en la Figura 4.11

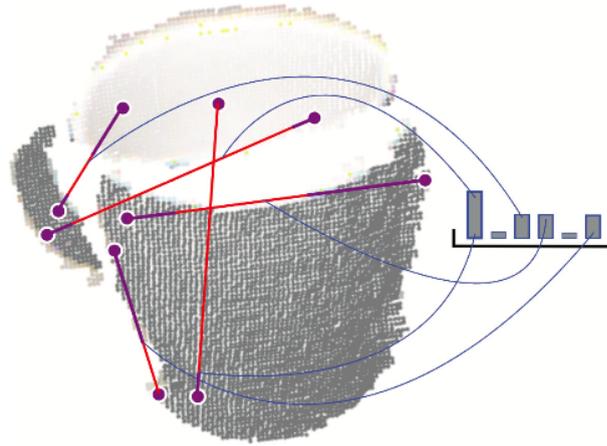


Figura 4.11: Ratio de líneas que yacen dentro y fuera de la superficie de la nube de ESF extraído de [99]

Los valores de otro grupo de variables se computa con la función de forma $A3$, codificando en un histograma los ángulos entre las dos líneas formadas por tres puntos al azar de la nube, como se observa gráficamente en la Figura 4.12.

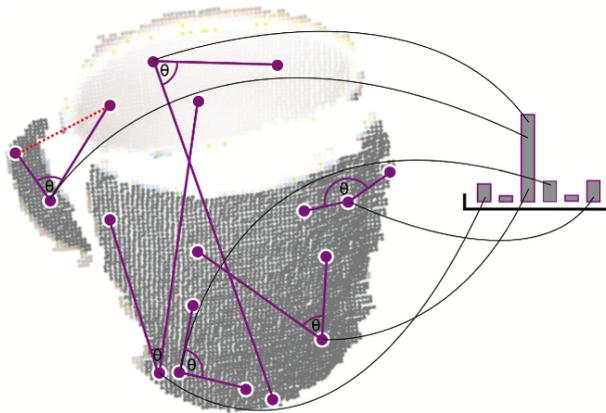


Figura 4.12: Función de forma $A3$ de ESF que cuantifica los ángulos obtenidos a partir de muestreo de tripletes, extraído de [99]

De forma análoga los ángulos se cuantifican contando la cantidad de ellos que quedan *en* la superficie de la nube, cuantos *fuera* y cuantos *parcialmente* en la superficie.

Finalmente se propone la función de forma $D3$ la cual cuantifica las áreas formadas por los tripletes muestreados, como se ve en la Figura 4.13. Además de forma análoga a las variables de distancias y líneas de pares de puntos, se estima otro grupo de variables que se construye codificando estas áreas y contando la cantidad de ellas que quedan *en* la superficie de la nube, cuantas *fuera* y cuantas *parcialmente* en la superficie.

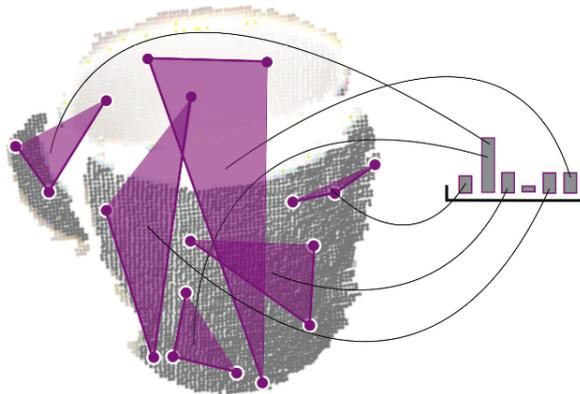


Figura 4.13: Función de forma $D3$ de ESF que cuantifica las áreas obtenidas a partir de muestreo de tripletes, extraído de [99]

4.4 Conclusiones

En este capítulo se presentaron las tecnologías de visión computacional y aprendizaje de máquinas supervisado evaluadas con el método propuesto en esta tesis. Como se puede observar y debido al carácter aplicado de la tesis, decidimos estudiar y evaluar tecnologías maduras y para las cuales existen implementaciones computacionales ya que el objetivo de la tesis es validar la existencia de correlaciones entre el aspecto visual y geométrico del brazo y su estado de activación muscular.

En el próximo capítulo presentamos una instancia del método que pretende lograr la estimación buscada a partir de un modelo aprendido para una cantidad considerable de sujetos de entrenamiento, pero generalizando a sujetos no pertenecientes a este conjunto de entrenamiento.

Capítulo 5

Estimación de activación muscular generalizando a nuevos sujetos

5.1 Introducción

El contenido de este capítulo ha sido publicado en [2] y aquí atacaremos el problema de estimar el nivel de activación discreto alcanzado por un brazo humano al levantar objetos de pesos variados en condiciones de contracción isométrica usando un modelo aprendido a partir de un gran conjunto de sujetos, del cual el sujeto a evaluar no forma parte. En este caso, la estimación se pretende lograr mapeando las características visuales de la piel extraídas con algoritmos de descripción de imágenes RGB aplicados a imágenes de la piel del brazo, con el nivel de activación correspondiente definido por el objeto sostenido. De esta manera se construye un modelo de clasificación de imágenes usando aprendizaje de máquinas supervisado. El conjunto de *ejemplos de entrenamiento* proveídos al algoritmo consisten en muchos pares de imágenes del bíceps de sujetos pertenecientes al conjunto de entrenamiento junto con la etiqueta que indica cual es el objeto que está siendo sostenido en esa imagen. El modelo aprendido a partir de este conjunto de entrenamiento es evaluado en la tarea de predecir el nivel de esfuerzo sobre el conjunto de *ejemplos de testeo*, el cual esta formado por imágenes del bíceps de sujetos no pertenecientes al conjunto de entrenamiento.

5.2 El enfoque

En esta sección explicaremos las especificaciones asumidas sobre el método general para este capítulo, gráficamente representado en la Figura 5.1.

Antes de entrar en el algoritmo, cada imagen es pre-procesada (Figura 5.1 (A)) eliminando de la imagen todo lo que no sea bíceps, es decir el antebrazo y el objeto, de forma manual y segmentando automáticamente el fondo, usando el algoritmo Gaussian Mixtures [102] pre entrenado con un conjunto de imágenes segmentadas manualmente.

En un segundo paso (Figura 5.1 (B)) se generan los vectores de características usados para entrenar los modelos usando algunos de los algoritmos de visión computacional previamente explicados en la Sección 4.3 para descripción de imágenes RGB: *Patrones binarios locales* (LBP) para extraer información de textura [69], e *Histogramas de color HS* (HSCCH del inglés Hue and Saturation Color Histogram) para extraer información de color [24, 93]. En ambas técnicas de descripción de imágenes, los tamaños de las grillas fueron validados previamente a través de un método que estimó el poder discriminador de cada tamaño de grilla, midiendo la entropía obtenida luego de agrupar los ejemplos por distancia en el espacio de características. El valor elegido para el tamaño de la grilla fue de 3 por 3 resultando en 9 regiones de interés. La razón principal de dividir las imágenes en grillas para generar las regiones de interés es poder mantener una correlación espacial de manera de analizar como la misma zona del brazo cambia para cada nivel de esfuerzo.

Con respecto a los algoritmos de aprendizaje supervisado aplicados para aprender el mapeo entre la imagen y el nivel de activación (Figura 5.1 (D)), consideramos en estos experimentos: *maquinas de vectores de soporte* (SVM) [30] y *bosques aleatorios* (RF) [17]. Para la técnica SVM evaluamos a su vez la performance de tres posibles kernels: *radial*, *linear* y *polynomial*. Los parámetros de este algoritmo fueron elegidos por

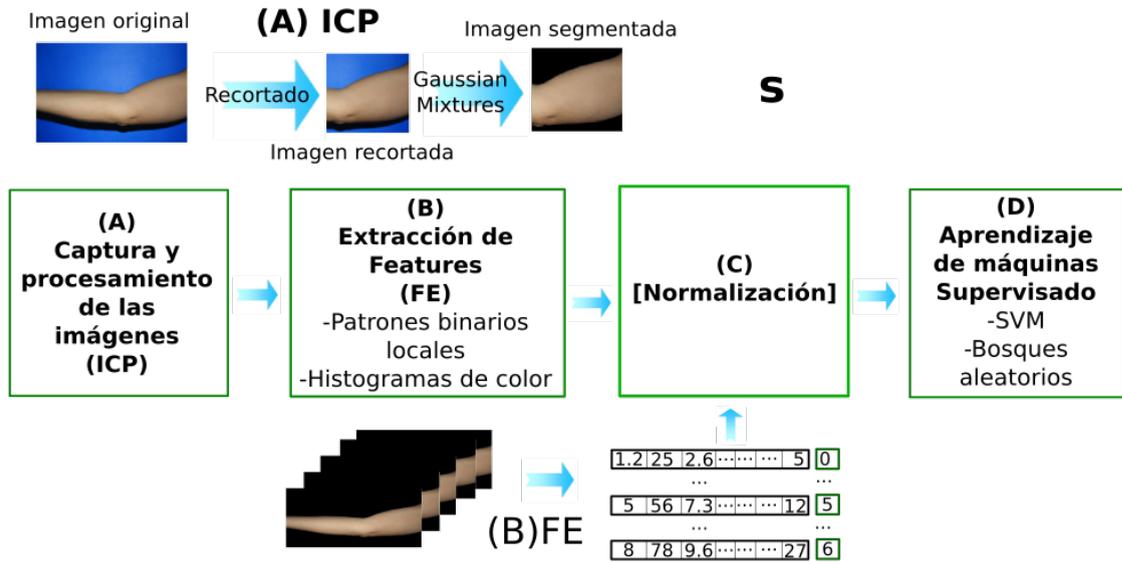


Figura 5.1: Enfoque utilizado

un proceso de tuning en el entrenamiento sobre una grilla de posibles valores usando 10 fold *cross-validation* sobre el conjunto de entrenamiento. Los parámetros y valores considerados para cada kernel fueron los siguientes:

- Para el kernel *radial*
 - C (el margen suave) fue elegido entre los valores $\{2 \times 10^4, 2 \times 10^5, \dots, 2 \times 10^{14}, 2 \times 10^{15}\}$
 - σ (la desviación estándar del kernel) fue elegida de los siguientes valores $\{2 \times 10^{-15}, 2 \times 10^{-14}, \dots, 2 \times 10^{-7}, 2 \times 10^{-6}\}$
- El kernel *linear* solo tiene el parámetro C cuyo valor se eligió de entre los siguientes $\{2 \times 10^{-15}, 2 \times 10^{-14}, \dots, 2 \times 10^{-2}, 2 \times 10^{-1}\}$
- Los parámetros del kernel *polynomial* fueron los siguientes:
 - C elegido entre los valores $\{2 \times 10^{-15}, 2 \times 10^{-14}, \dots, 2 \times 10^{-2}, 2 \times 10^{-1}\}$
 - El *grado* del kernel *polynomial* elegido de los valores $\{1, 2, 3, 4\}$
 - La *escala* del kernel *polynomial* elegida de los valores: $\{1, 2, 3\}$

Estos rangos fueron elegidos de experimentación preliminar previa sobre rangos mas amplios. Los bosques aleatorios fueron optimizados sobre la cantidad de variables usadas para construir cada uno de los árboles, considerando diferentes proporciones del total del número de variables del vector de entrada: $\{2\%, 6\%, 12\%, 25\%, 50\%, 75\%, 100\%\}$.

Antes de concluir esta sección presentamos una variación del enfoque que ayuda a mejorar los resultados en la tarea de generalizar a sujetos no antes vistos, aunque toma más información que el enfoque estándar. Para la clasificación de una imagen en su nivel de activación, esta mejora no solo considera como entrada la imagen a evaluar si no que además considera la imagen del bíceps para el sujeto en evaluación para el caso de activación mínima. Aunque esta ayuda rompe con el protocolo clásico de clasificación, es algo que puede ser fácilmente logrado en la práctica y lo nombramos con el concepto de *calibración*. Esta calibración se realiza *normalizando* (Figura 5.1 (C)) los vectores de features, restando de cada uno el vector de features correspondiente a la mínima activación para el sujeto en análisis. El proceso de *normalización* se ejemplifica gráficamente en la Figura 5.2 donde el vector $FVSiO1k$ es el vector de features original del sujeto i para la clase $O1k$, obtenido aplicando alguna de las técnicas de descripción de imágenes presentadas, a la imagen que se desea evaluar (que en este ejemplo corresponde a un nivel de activación de $O1k$). El vector $FVSiO0$ es el vector de features del sujeto i para la clase de mínimo nivel de activación $O0$, obtenido aplicando la misma

técnica de descripción de imágenes a una imagen correspondiente a este mínimo nivel de activación. Como se ve en la figura, este último se le resta al primero y se obtiene el vector normalizado llamado *NormFVSiO1k* en esta figura, el cual es etiquetado para este ejemplo con la clase *O1k*. La motivación de esta variación esta basada en el hecho de que generalizar a sujetos no antes vistos es una tarea muy compleja como se verá en los resultados experimentales de la Sección 5.3. Esta dificultad se debe a que los brazos pueden diferir en muchos aspectos que no tienen relación con la contracción muscular, lo que agrega mucho ruido al problema. En particular, los brazos pueden producir patrones en la piel muy diferentes para el mismo nivel de activación para sujetos distintos. Al observar los experimentos veremos que es demasiado pedir esta generalización sin ningún tipo de información previa del sujeto y veremos también como esta calibración ayuda en el proceso, al costo de proveer al menos ejemplos etiquetados para el menor nivel de contracción.

$$\begin{array}{r}
 \text{FVSiO1k} \quad \boxed{5 \mid 56 \mid 7.3 \mid \dots \mid \dots \mid 12} \\
 - \\
 \text{FVSiO0} \quad \boxed{1.2 \mid 25 \mid 2.6 \mid \dots \mid \dots \mid 5} \\
 = \\
 \text{NormFVSiO1k} \quad \boxed{3.8 \mid 31 \mid 4.7 \mid \dots \mid \dots \mid 7} \quad \boxed{\text{O1k}}
 \end{array}$$

Figura 5.2: Proceso de normalización de vectores de características

5.3 Evaluación empírica

5.3.1 Setup experimental

Evaluamos el enfoque para distintos escenarios: la tarea de estimar 2 y 4 niveles de activación discretos; las técnicas *LBP* o *HSCCH*; los algoritmos *SVM* o *RF* con las tres posibles variantes de *SVM* eligiendo entre los kernels *linear*, *radial*, o *polynomial*; y consideramos para cada caso el uso o no de calibración por *normalización (N)*.

El problema atacado en esta tesis es novedoso y complejo, lo que resulta en que no hayan enfoques competidores que resuelvan el mismo problema con quien comparar. Resultando en el clasificador aleatorio como un competidor válido ya que consideramos valioso encontrar aunque sea una mínima generalización entre sujetos.

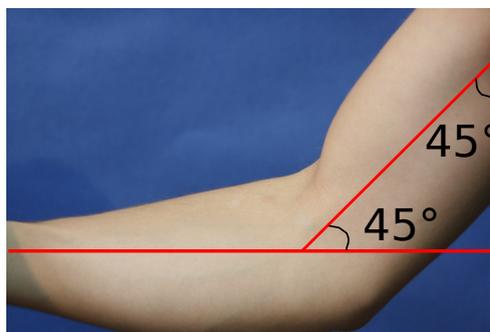


Figura 5.3: Posición de captura

Para estos experimentos capturamos imágenes RGB de 100 sujetos, sosteniendo cada uno 5 objetos posibles de peso 0 g, 500 g, 1300 g, 1900 g y 3800 g (llamados de aquí en adelante con las etiquetas *O0*, *O500*, *O1k*, *O2k* y *O4k* respectivamente), y 3 imágenes por cada uno de los pesos. Un ejemplo de una imagen cruda del corpus se muestra en la Figura 5.3. Todas las imágenes fueron capturadas usando una cámara de fotos reflex Canon EOS 1000D con Zoom de 75-300 mm, sin flash y con foco automático, tomadas lo mas cerca del bíceps posible e iluminadas artificialmente con un reflector de lampara incandescente de 60

W, ubicado aproximadamente 45° a la izquierda y al frente del sujeto, elevado aproximadamente 1.5 mts del sujeto y apuntando al brazo. Además, todas estas imágenes fueron capturadas con un fondo azul para facilitar el proceso de segmentación automática.

Para asegurar variabilidad entre los 100 sujetos, incluimos tanto hombres como mujeres; una diversidad de colores de piel que excluye colores de piel típicos de personas africanas y asiáticas; y sujetos con diversos niveles de actividad física desde intenso a moderado. Además consideramos un amplio rango de edades entre 17 y 54 años, y un amplio rango de pesos entre 47 y 113 kg, tanto como un amplio rango de alturas variando entre los 1.6 mts y los 1.9 mts, lo que resulto en un rango de índice de masa corporal entre 19.32 y 34.16. Los valores máximos, mínimos, promedio y la desviación estándar de la edad, peso, altura e índice de masa corporal de los sujetos esta resumida en la Tabla 5.1.

	Max	Min	Mean	SD
Edad	54,00	17,00	22,55	5,27
Peso [Kgs]	113,00	47,00	73,17	12,27
Altura [mts]	1,93	1,55	1,74	0,08
Indice de masa corporal	37,76	18,79	24,06	3,57

Tabla 5.1: Resumen de información de los sujetos

Para el caso donde se realiza *normalización*, el objeto de 0 g (clase *O0*, mínimo nivel de activación) no se considera para la predicción ni para el entrenamiento ya que su información es usada para modificar los ejemplos de entrenamiento, considerando entonces clasificación sobre las demás clases *O500* hasta *O4k*. Debido a esto es que restringimos los casos no normalizados también a 4 clases para permitir una comparación válida entre los resultados de ambos escenarios.

El procedimiento mencionado anteriormente produce un total de 1500 imágenes que, luego de aplicar los algoritmos de extracción de features visuales (ya sea *LBP* o *HSCCH*), y etiquetar cada ejemplo con la clase correspondiente al nivel de activación ejercido, produce el dataset etiquetado. Estos datasets son usados con el enfoque estándar de testeo, donde son particionados en conjuntos disjuntos de entrenamiento y testeo, con el primero usado como entrada a los algoritmos de clasificación (ya sea *SVM* o *Random Forests*) para entrenar los modelos, y el ultimo usado para evaluar el error de generalización de los modelos aprendidos. Es importante recalcar que estas divisiones de los conjuntos de entrenamiento y testeo mantienen todas las imágenes de un mismo sujeto en la misma partición, garantizando que la generalización se evalúa sobre sujetos no vistos en entrenamiento.

Para un mejor reporte de resultados y a la misma vez tomar ventaja de los datos limitados se tomaron 10 pares de conjuntos de entrenamiento/testeo. Cada par fue generado tomando al azar 10 de los 100 sujetos y produciendo como conjunto de testeo las imágenes correspondientes a estos 10 sujetos (resultando en 120 imágenes para el problema de 4 clases y 60 para el de 2 clases); el conjunto de entrenamiento se construyo con las imágenes de los restantes 90 sujetos (resultando en 1080 imágenes para el problema de 4 clases y 540 para el problema de 2 clases)

En este trabajo reportamos el error de generalización de los modelos usando las medidas de performance para clasificación previamente presentadas: *accuracy*, *precision*, *recall*, *f-measure* y *MAE*. Cada una de estas medidas se computa para cada uno de las 10 particiones y se reporta el promedio y la desviación estándar.

5.3.2 Resultados experimentales

En esta sección presentamos los resultados de los experimentos realizados.

Estimación de activación muscular

Como se explicó anteriormente, corrimos todas las posibles variaciones del método presentado en este capítulo resultando en:

1. la elección del algoritmo de aprendizaje: ya sea *SVM* con kernel radial (*RadialSVM*), *SVM* con kernel linear (*LinearSVM*), *SVM* con kernel polynomial (*PolySVM*), Bosques Aleatorios (*RF*), o el clasificador aleatorio como competidor de base (*RDM*)

ALG	D	N	\hat{P}	\hat{R}	$\hat{F}\hat{M}$	\hat{A}	OvA	MAE
RadialSVM	RDM		54.09(5.95)	54(5.83)	53.8(5.81)	54(5.83)	54(5.83)	1518(192.63)
		HSCH	FALSE	64.27(9.63)	63.66(9.22)	63.28(9.31)	63.66(9.22)	63.66(9.22)
	LBP	TRUE	70.52(6.04)	69.66(6.37)	69.15(7.00)	69.66(6.37)	69.66(6.37)	1001(210.31)
		FALSE	75.78(7.76)	75.44(7.83)	75.34(7.88)	75.44(7.83)	75.44(7.83)	810.27(258.59)
		TRUE	80.69(4.58)	79.60(5.19)	79.37(5.43)	79.61(5.16)	79.61(5.16)	672.58(170.45)
		FALSE	68.91(9.9)	67.83(9.23)	67.45(9.24)	67.83(9.23)	67.83(9.23)	1061.5(304.63)
LinearSVM	HSCH	TRUE	70.56(5.81)	69.66(5.97)	69.23(6.33)	69.66(5.97)	69.66(5.97)	1001(197.11)
		FALSE	76.68(4.67)	75.94(4.42)	75.79(4.40)	75.94(4.42)	75.94(4.42)	793.77(146.04)
	LBP	TRUE	80.59(6.73)	79.77(6.99)	79.60(7.13)	79.78(6.97)	79.78(6.97)	667.08(230.29)
		FALSE	69.01(9.40)	68.33(8.81)	68.05(8.95)	68.33(8.81)	68.33(8.81)	1045(291.03)
		TRUE	70.63(5.84)	69.66(5.86)	69.22(6.20)	69.66(5.86)	69.66(5.86)	1001(193.67)
		FALSE	77.16(5.27)	76.28(5.05)	76.09(5.05)	76.27(5.05)	76.27(5.05)	782.77(166.80)
PolySVM	HSCH	TRUE	79.83(6.41)	78.93(6.75)	78.74(6.89)	78.95(6.73)	78.95(6.73)	694.58(222.28)
		FALSE	68.37(6.85)	67.16(7.07)	66.37(7.80)	67.16(7.07)	67.16(7.07)	1083.5(233.41)
	LBP	TRUE	73.28(7.26)	71.83(6.68)	71.43(6.69)	71.83(6.68)	71.83(6.68)	929.5(220.68)
		FALSE	77.20(8.03)	75.94(7.40)	75.69(7.47)	75.93(7.43)	75.93(7.43)	794.14(245.44)
		TRUE	75.53(7.39)	73.74(6.68)	73.35(6.74)	73.78(6.67)	73.78(6.67)	864.99(220.25)
		FALSE	68.37(6.85)	67.16(7.07)	66.37(7.80)	67.16(7.07)	67.16(7.07)	1083.5(233.41)
RF	HSCH	TRUE	73.28(7.26)	71.83(6.68)	71.43(6.69)	71.83(6.68)	71.83(6.68)	929.5(220.68)
		FALSE	77.20(8.03)	75.94(7.40)	75.69(7.47)	75.93(7.43)	75.93(7.43)	794.14(245.44)
	LBP	TRUE	75.53(7.39)	73.74(6.68)	73.35(6.74)	73.78(6.67)	73.78(6.67)	864.99(220.25)
		FALSE	68.37(6.85)	67.16(7.07)	66.37(7.80)	67.16(7.07)	67.16(7.07)	1083.5(233.41)
		TRUE	73.28(7.26)	71.83(6.68)	71.43(6.69)	71.83(6.68)	71.83(6.68)	929.5(220.68)
		FALSE	77.20(8.03)	75.94(7.40)	75.69(7.47)	75.93(7.43)	75.93(7.43)	794.14(245.44)

Tabla 5.2: Resultados de la estimación para 2 niveles de activación: $O500$ y $O4k$. Los casos con normalización están resaltados en gris y las mejores performance se muestran en negritas. La columna 'ALG' describe al algoritmo de aprendizaje usado, la columna 'D' describe la técnica de extracción de características, y la columna 'N' indica si la normalización fue aplicada.

- el número de clases objetivo: 2 ($O500$ y $O4k$) o 4 ($O500$, $O1k$, $O2k$ y $O4k$)
- la técnica de descripción visual: patrones binarios locales (LBP), o los histogramas de color HS ($HSCH$)
- la elección de usar *normalización* o no

Los resultados se muestran en dos tablas, con la Tabla 5.2 mostrando los resultados de la clasificación en 2 niveles de activación y la Tabla 5.3 mostrando los resultados de la clasificación en 4 niveles de activación. Cada tabla muestra los resultados para todos los casos con las columnas 'ALG', 'D', y 'N' indicando la elección del algoritmo de aprendizaje, la técnica de extracción de características y el uso o no de normalización respectivamente. Las 6 columnas restantes reportan el *promedio* y la *desviación estándar* (entre paréntesis) de las medidas de performance elegidas (\hat{P} , \hat{R} , $\hat{F}\hat{M}$, \hat{A} , OvA , y MAE) sobre los 10 train-test datasets generados al azar de la división aleatoria de los 100 sujetos.

En el problema más simple de estimar los dos niveles de activación mas lejanos (500 g a 3800 g), los resultados de la Tabla 5.2 prueban como la *normalización* es útil ya que en todos los casos salvo en el $RF-LBP$, hay una mejora en las medidas de performance. A pesar de que ninguno de ellos es estadísticamente significativo, la tendencia es clara en favor de la *normalización*. La mayor mejora sucede para el caso $LinearSVM-LBP$ donde las medidas de clasificación mejoran casi un 5% y el MAE disminuye más de 130 g de 793 g a 667 g. Vale la pena también mencionar las mejoras contra el competidor de base, el clasificador aleatorio, donde en el mejor caso de $LinearSVM-LBP$, las medidas de clasificación mejoran cerca del 25% desde valores en el orden de 55% para el clasificador aleatorio contra valores en el orden de 80% para el caso $LinearSVM-LBP$ y el MAE disminuye de 1500 g a 667 g.

Tendencias similares se ven en la Tabla 5.3 que muestra los resultados del caso de estimación de 4 niveles de activación ($O500$, $O1k$, $O2k$ y $O4k$). Este caso es mas desafiante tanto por los valores extremos que son los mismos que en el caso de 2 clases (500 g y 3800 g) y se agregan dos valores intermedios (1300 g y 1900 g), lo que resulta en una reducción de la discriminación visual entre ejemplos. En primer lugar, notamos una esperada disminución en todas las medidas de performance para todos los casos, incluido el clasificador aleatorio que reduce sus valores desde cerca del 50% para el caso binario al orden de 25% para el caso de 4 clases. Sin embargo, el MAE del clasificador aleatorio mejora (reduciéndose a 1295 g) ya que la disminución en los errores de clasificación viene acompañada por menores diferencias en gramos entre las clases. A pesar de esto se ven tendencias similares al caso binario. En primer lugar notamos que la mejor performance (mostrada en negritas) se obtiene para la misma técnica de extracción de features LBP , y el mismo algoritmo de aprendizaje SVM , aunque con un kernel radial en lugar de kernel lineal. También notamos que la normalización ayuda nuevamente, generando una mejora en el mejor caso de 5% para el caso de $RadialSVM-LBP$. Esta vez las mejoras suceden para todos los casos, incluyendo RF con LBP (que no tuvo mejoras en el caso binario). También se ven medidas de performance de clasificación superiores casi un 20% sobre el random (para el caso de $RadialSVM-LBP$).

ALG	D	N	\hat{P}	\hat{R}	$\hat{F}M$	\hat{A}	OvA	MAE
RadialSVM	RDM		24.79(3.39)	24.87(3.28)	24.68(3.31)	62.43(1.64)	24.87(3.28)	1295.47(83.52)
	HSCH	FALSE	34.29(4.07)	34.85(3.89)	33.51(3.64)	67.44(1.97)	34.89(3.95)	1093.56(165.99)
		TRUE	39.39(5.89)	39.46(5.81)	38.13(5.48)	69.74(2.90)	39.48(5.81)	990.09(75.78)
	LBP	FALSE	38.33(5.20)	36.74(5.26)	36.36(5.44)	68.37(2.62)	36.75(5.25)	947.88(111.25)
		TRUE	43.32(4.82)	42.32(4.43)	41.63(4.68)	71.17(2.19)	42.34(4.39)	832.04(69.29)
	LinearSVM	HSCH	FALSE	34.02(2.59)	34.44(2.52)	32.96(2.86)	67.23(1.27)	34.47(2.55)
TRUE			37.21(6.17)	37.47(5.79)	36.24(6.01)	68.73(2.89)	37.47(5.79)	1019.23(82.50)
LBP		FALSE	36.47(4.37)	36.23(4.63)	35.62(4.59)	68.13(2.32)	36.26(4.64)	905.93(66.21)
		TRUE	41.53(4.21)	40.56(4.45)	39.86(5.01)	70.29(2.20)	40.59(4.40)	838.54(81.16)
PolySVM	HSCH	FALSE	32.73(3.84)	32.78(3.68)	31.87(3.73)	66.40(1.85)	32.80(3.7)	1090.87(122.79)
		TRUE	37.56(6.19)	37.63(5.82)	36.54(6)	68.82(2.91)	37.64(5.83)	1011.33(77.5)
	LBP	FALSE	37.36(3.25)	36.73(3.43)	36.22(3.29)	68.37(1.71)	36.75(3.42)	902.10(53.53)
		TRUE	41.54(4.94)	40.65(4.5)	39.90(5.11)	70.33(2.23)	40.67(4.46)	836.23(84.08)
RF	HSCH	FALSE	30.67(3.46)	31.78(2.97)	29.86(2.99)	65.9(1.48)	31.8(2.97)	1139.68(109.13)
		TRUE	35.06(3.95)	35.36(4.66)	33.80(4.59)	67.69(2.32)	35.39(4.65)	1067.41(100.75)
	LBP	FALSE	30.54(8.14)	32.46(5.52)	30.45(6.34)	66.25(2.77)	32.5(5.54)	1041.66(100.67)
		TRUE	35.51(4.77)	37.39(4.27)	35.39(4.02)	68.71(2.13)	37.42(4.27)	956.21(97.73)

Tabla 5.3: Resultados de la estimación para 4 niveles de activación: $O500$, $O1k$, $O2k$ y $O4k$. Los casos con normalización están resaltados en gris y las mejores performance se muestran en negritas. La columna 'ALG' describe al algoritmo de aprendizaje usado, la columna 'D' describe la técnica de extracción de características, y la columna 'N' indica si la normalización fue aplicada.

Curvas de aprendizaje

Presentamos en esta sección los resultados de las *curvas de aprendizaje* [75, 6], explicadas en la Sección 4.2.4, para los modelos entrenados. Las mismas sugieren que posibles mejoras pueden obtenerse entrenando con más de 90 sujetos.

Presentamos las curvas de aprendizaje de: el problema de dos clases con clasificador SVM Linear en la Figura 5.4, el problema de dos clases con bosques aleatorios como clasificador en la Figura 5.5, el problema de cuatro clases con el clasificador SVM Radial en la Figura 5.6, y finalmente el problema de cuatro clases con el clasificador de bosques aleatorios en la Figura 5.7. La razón de estas elecciones fue en primer lugar elegir los casos ganadores para los problemas de dos y cuatro clases (SVM Linear y SVM Radial respectivamente), y agregar el caso de bosques aleatorios por diversidad. Además se considero solo el caso base, sin normalización ya que este caso representa el ideal en el que no se necesita información extra para la predicción.

Para construir las curvas, primero dividimos el conjunto de 100 sujetos en grupos de 90/10 para generar un training set con los ejemplos etiquetados correspondientes al grupo de 90 sujetos y un conjunto de testeo con los ejemplos etiquetados correspondientes a los otros 10. Repetimos el proceso 3 veces para generar 3 pares de conjuntos de entrenamiento y testeo. Luego muestreamos conjuntos de entrenamiento de tamaños menores al total $m = \{50, 60, 70, 80, 90\}$, con 10 muestras por m (con la excepción de $m = 90$ que produce siempre el mismo sample). Posteriormente entrenamos un clasificador para cada una de estas 10 repeticiones de m y reportamos el promedio y la desviación estándar del error de generalización sobre el test-set y sobre el train-set. Para hacer una comparación válida se mantuvo el mismo test-set para los mismos $4 \times 10 + 1$ casos. Para estas curvas se reporta el MAE como la medida más representativa, asumiendo que las conclusiones sobre esta medida no van cambiar para las demás medidas. La media y desviación estándar del MAE sobre las tres divisiones 90/10 se muestran en las figuras como puntos y barras verticales respectivamente con el error de generalización con cuadrados grises y el de entrenamiento con diamantes negros.

Todas las figuras muestran un comportamiento similar para los errores de generalización (test-set). Estos errores son altos y no muestran una tendencia a disminuir estadísticamente significativa. Por otro lado, las curvas del error de entrenamiento de casi todas las figuras se mantienen muy cercanas a un MAE de 0 g para todos los m , con la excepción del caso de SVM Radial sobre cuatro clases (Figura 5.6) que muestra un valor promedio cerca de los 200 g y una tendencia creciente con m . Sin embargo, en todos los casos el error de entrenamiento no ha alcanzado el error de generalización. Como se discutió anteriormente, los dos errores deberían converger al mismo valor, por lo que si bien no podemos garantizar que se encontraran en errores menores, esa posibilidad no se descarta de estas curvas. Estas curvas demuestran que hay posibilidad de mejora en los resultados si aumentamos la cantidad de sujetos usados para entrenar los modelos.

Finalizamos el análisis de las curvas de aprendizaje con la Tabla 5.4 que reproduce los resultados de los errores de generalización y entrenamiento para el caso de $m = 90$, pero con 10 repeticiones en vez de 3 como en las figuras. De la presentación anterior debe quedar claro que la distancia de cero del error de entrenamiento junto con el gap entre el error de entrenamiento y el de generalización provee toda la

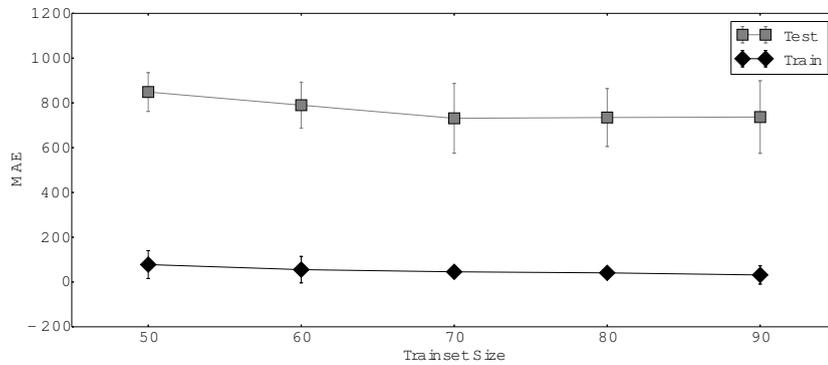


Figura 5.4: Curvas de aprendizaje para LBP con SVM Linear, sin normalización y clasificación en dos clases ($O500$ y $O4k$). La curva de diamantes negros reporta el MAE sobre el train-set y la curva de cuadrados grises reporta el MAE sobre el test-set. El eje horizontal reporta m , el número de sujetos considerados en el conjunto de entrenamiento usado para aprender el clasificador.

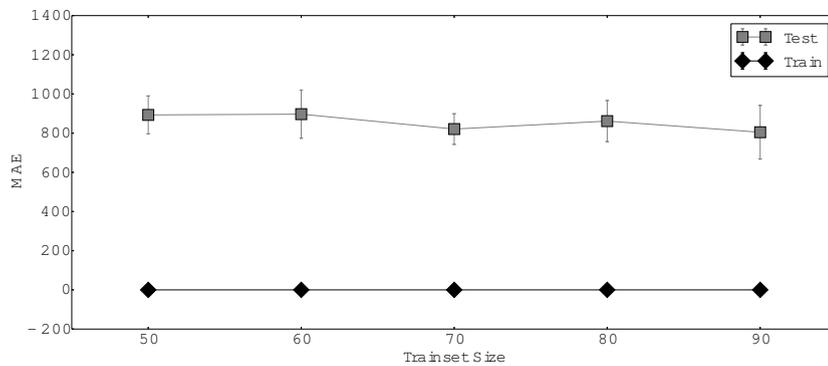


Figura 5.5: Curvas de aprendizaje para LBP con busques aleatorios, sin normalización y clasificación sobre 2 clases ($O500$ y $O4k$). La curva de diamantes negros reporta el MAE sobre el train-set y la curva de cuadrados grises reporta el MAE sobre el test-set. En el eje horizontal se reporta m , el número de sujetos usados para entrenar el clasificador

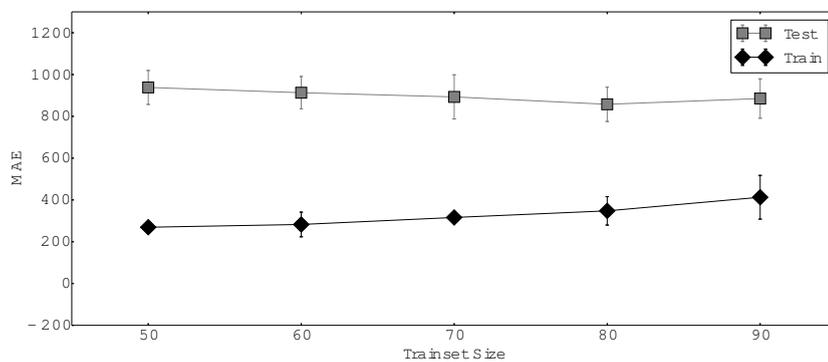


Figura 5.6: Curvas de aprendizaje para LBP con SVM Radial, sin normalización y clasificación en cuatro clases ($O500$, $O1k$, $O2k$ y $O4k$). La curva de diamantes negros reporta el MAE sobre el train-set y la curva de cuadrados grises reporta el MAE sobre el test-set. En el eje horizontal se reporta m , el número de sujetos usados para entrenar el clasificador

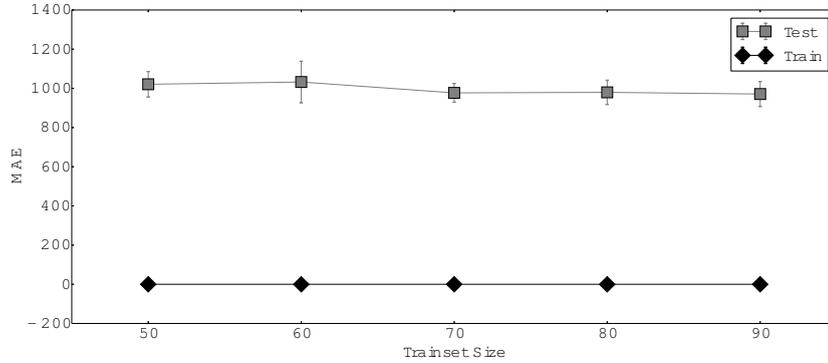


Figura 5.7: Curvas de aprendizaje para LBP con bosques aleatorios, sin normalización y clasificación en cuatro clases ($O500$, $O1k$, $O2k$ y $O4k$). La curva de diamantes negros reporta el MAE sobre el train-set y la curva de cuadrados grises reporta el MAE sobre el test-set. En el eje horizontal se reporta m , el número de sujetos usados para entrenar el clasificador

N° Classes	ALG	TeMAE	TrMAE
2	LinearSVM	705.67(257.23)	75.3(94.23)
	RF	793.68(333.38)	0.00(0)
4	RadialSVM	924.61(161.11)	361.62(288.81)
	RF	973.81(170.29)	0.00(0)

Tabla 5.4: Resultados de la estimación sobre el train-set y el test-set para dos y cuatro niveles de activación usando LBP como descriptor y sin normalización. La columna 'ALG' describe el algoritmo usado y las columnas 'TeMAE' y 'TrMAE' reportan el MAE de entrenamiento y de generalización respectivamente.

información necesaria para llegar a la conclusión de que entrenar los modelos con más datos puede mejorar (bajar) el error de generalización. Por esa razón decidimos correr los experimentos para el caso especial de $m = 90$ pero con un soporte estadístico mayor reportando el promedio de los dos errores para 10 repeticiones, cada una con un conjunto diferente y no superpuesto de conjunto de testeo generado del conjunto original de 100 sujetos. La tabla muestra resultados similares, con conclusiones similares a aquellos obtenidos por las figuras. Casi todos los casos muestran un error de entrenamiento de cero (considerando la desviación estándar) con la excepción del caso de Radial SVM y cuatro clases que muestra un MAE menos la desviación estándar de apenas 73 gr. Además, los cuatro casos muestran un gran gap entre los dos errores, mostrando que hay todavía gran lugar para mejorar el error de generalización.

5.4 Discusión y conclusiones

De estos resultados podemos afirmar que es posible diferenciar el nivel de activación muscular que una persona esta ejerciendo a partir de una foto de un bíceps: entre 4 niveles con una FMeasure $\approx 41\%$ y Accuracy $\approx 71\%$; y entre 2 niveles con una FMeasure $\approx 79\%$ y Accuracy $\approx 79\%$. Además es posible estimar el peso que una persona esta sosteniendo con su brazo a partir de una foto de su bíceps: entre 4 pesos diferentes (500 g, 1300 g, 1900 g y 3800 g) con un error absoluto medio ≈ 832 g y entre 2 pesos diferentes (500 g y 3800 g) con un error absoluto medio de ≈ 667 g. Como puede observarse, para el caso de 4 pesos, la diferencia promedio entre los niveles de activación contiguos es de aproximadamente 800 g en promedio, similar al error obtenido; esto valida el hecho de que los errores cometidos son entre niveles contiguos de activación.

Estos resultados experimentales prueban que nuestro enfoque es capaz de encontrar patrones discriminatorios entre distintos niveles de activación indicando no solo la existencia de correlaciones entre la activación muscular e imágenes del bíceps, si no también que estas correlaciones son extraíbles mediante sistemas autónomos. Aunque este es un paso hacia una aplicación práctica en el futuro, la precisión alcanzada no esta cerca de un uso práctico y en los mejores casos de normalización se usa información etiquetada del

sujeto a medir.

La primera contribución de este capítulo es la experimentación de este enfoque sobre un conjunto de datos de imágenes de 100 sujetos realizando 4 y 2 diferentes niveles discretos de esfuerzo en postura estática de contracción isométrica, siendo estos los primeros pasos en esta línea de investigación. En este capítulo, proponemos evaluar *histogramas de color HS* y *patrones binarios locales* como descriptores de imagen. Las técnicas de aprendizaje automático que se evaluaron en este capítulo fueron *maquinas de vectores de soporte* y *bosques aleatorios*. Un hallazgo importante de este estudio es que el principal problema con este enfoque es la dificultad para generalizar el modelo sobre sujetos no vistos, debido a las grandes diferencias existentes entre los brazos de sujetos humanos, requiriendo potencialmente imágenes para una gran cantidad de sujetos (muchos más de 100) para aprender un modelo adecuado. Para mejorar los resultados de la estimación, ayudando al enfoque a generalizar mejor sobre sujetos que no se ha visto anteriormente, otra contribución es la propuesta de *normalización* de los vectores de características en relación con los vectores de características de activación mínimo como forma de calibración. Sin embargo, estos resultados no son suficientes para una aplicación práctica, que es una de las limitaciones de la variación del método presentada en este capítulo. Otra limitación es el enfoque simplista considerado, especialmente el uso de algoritmos de extracción de características de uso general que usan solo información de intensidad de píxeles para inferir el estado muscular, sin considerar información 3D más informativa. Finalmente, la necesidad de imágenes sobre la piel expuesta también es un factor limitante.

Siguiendo la intuición de que la información específica del sujeto es necesaria para una buena estimación, en el siguiente capítulo se considera un setting más sencillo donde la generalización pretendida es intra-sujeto. En primer lugar se evalúa el uso de imágenes RGB para este problema en la Sección 6.2. Posteriormente en la Sección 6.3 se considera una aplicación práctica en tele-fisioterapia donde la necesidad de entrenar los modelos para cada sujeto en particular no es un limitante. Además en esa misma sección, para superar la limitación que impone la información RGB, como información de entrada se considera el uso de nubes de puntos 3D que proveen más información sobre el estado de activación del músculo materializada en la deformación del brazo.

Capítulo 6

Estimación de activación muscular a partir de modelos especializados para cada sujeto

6.1 Introducción

Considerando los resultados obtenidos en la experimentación presentada en el Capítulo 5, donde se comprueba la complejidad del problema de estimar la activación muscular sin ningún tipo de información previa del sujeto a evaluar, es que en este capítulo se propone relajar esa restricción. En este capítulo se ataca el problema de estimar el nivel de activación muscular para un sujeto, a partir de un modelo aprendido con información específica de este mismo sujeto. Se acepta la posibilidad de relajar la restricción de generalización en sujetos debido a que existen algunas aplicaciones en las que es posible aprender o calibrar el modelo para un sujeto en particular y luego usarlo en otras circunstancias. Un ejemplo de esa posible aplicación es en tele-medicina como se la presenta en la Sección 6.3. Además presentamos en este capítulo una evaluación preliminar que intenta validar el método, instanciado usando imágenes RGB como entrada, en la Sección 6.2.

6.2 Evaluaciones preliminares usando imágenes RGB como entrada

6.2.1 El problema

En esta sección se presenta una evaluación preliminar del método propuesto en esta tesis, en la tarea de estimación de activación muscular de un sujeto específico a partir de un modelo personalizado para el mismo. El contenido de esta sección ha sido publicado en [1]. El problema específico se restringe al uso de imágenes RGB como entrada para el método. Además se conservan las mismas condiciones de escala, iluminación y punto de vista para todas las imágenes usadas tanto para entrenar el modelo como para evaluarlo.

6.2.2 El enfoque

Para producir los ejemplos de mapeo, se capturaron imágenes RGB del brazo (usando una cámara económica del tipo reflex) desde un punto de vista fijo, mientras que la mano correspondiente sostenía objetos de pesos conocidos. Esos valores de peso se tomaron como medida del nivel de activación del bíceps.

Luego de capturar las imágenes es necesario un realizar la segmentación del brazo. Para ello se propone un proceso semi-automático de segmentación que consiste en identificar en la imagen los píxeles que son del color del fondo (azul) y eliminarlo como background. La segmentación se realizó aplicando el algoritmo de clustering K-Medias [27] tomando como data-points los valores RGB de cada píxel de las imágenes. Posteriormente se aplicó una operación morfológica de erosión para remover posible ruido. Finalmente a los píxeles del background en la imagen original se les asignó el valor cero.

Para extraer un vector de características de las imágenes segmentadas de manera de poder usarlo como entrada en el proceso de aprendizaje, las técnicas que se evaluaron fueron *bolsa de puntos clave (BKP)*, *patrones binarios locales (LBP)*, *histograma de color HS (HC)* y *momentos de contornos (MC)* (todas presentadas en la Sección 4.3.1). Como técnica de aprendizaje de máquinas se evaluó en esta sección el algoritmo *SVM*.

6.2.3 Evaluación empírica

Setup experimental

Para capturar las imágenes se procede de la siguiente manera:

1. El sujeto de prueba sostiene una carga de peso P_i con el brazo estirado, a piel descubierta
2. Se toma una fotografía solamente del brazo (sin capturar el objeto) con un fondo azul para facilitar posteriormente el proceso de segmentación del brazo
3. Para un P_i dado se repiten los pasos 1 y 2, N veces para obtener N ejemplos del ese nivel de carga
4. Se repite el proceso anterior con el resto de los pesos.

Como resultado del proceso de segmentación y extracción de características previamente presentados, se obtuvieron un dataset por cada técnica. Los vectores de características de cada imagen se etiquetaron con el objeto levantado. De esta forma cada dataset contiene un data-point por imagen que consiste en una tupla formada por el vector de características extraídas para la imagen, y la clasificación correcta de la imagen.

Realizamos el proceso de aprendizaje y clasificación sobre un conjunto de 92 imágenes segmentadas, de brazos sosteniendo 3 pesos distintos. Los pesos y objetos utilizados son: el objeto O_0 que pesa 0 Kg, el objeto O_5 que pesa 2.25 Kg y el objeto O_6 que pesa 5 Kg. Estos pesos corresponden con las clases 0, 5 y 6 respectivamente. Se tienen 30 instancias de la clase 0, 33 instancias de la clase 5 y 29 instancias de la clase 6. Cada dataset se fragmentó en un conjunto de entrenamiento conformado por el 70% del dataset original y el conjunto de testeo conformado por el 30% restante tomadas al azar del dataset completo. En distintas instancias del experimento se entrenó el clasificador usando datos extraídos con los 4 tipos de características presentadas, se reportaron las medidas de rendimiento presentadas anteriormente y se realizó un análisis comparativo de las mismas.

En esta sección evaluamos la técnica de aprendizaje de máquinas *máquinas de vectores de soporte* [30] y utilizamos los valores por defecto de la función implementada con el paquete *e1071* del lenguaje de programación R [65] para sus parámetros, siendo los más importantes un kernel de tipo radial con $\gamma = 0.0018$. Usando estos recursos aprendimos el modelo de los datos subyacentes sobre el dataset de entrenamiento.

Resultados experimentales

Como medidas de rendimiento se consideraron *precisión* (\hat{P}), *recall* (\hat{R}), *F-Measure* ($\hat{F}\hat{M}$) y *accuracy* (\hat{A}), para el caso multi-clase para cada experimento junto con la *overall accuracy* (OvA).

Resumimos en la Tabla 6.1 los valores promedio y de OvA obtenidos para cada medida de rendimiento sobre cada dataset obtenido con las distintas técnicas de extracción de características.

	OvA	\hat{A}	\hat{P}	\hat{R}	$\hat{F}\hat{M}$
BKP	82.14%	88.09%	80.83%	82.78%	80.83%
LBP	100%	100%	100%	100%	100%
HC	75%	83.33%	76.06%	73.56%	73.29%
MC	92.85%	95.23%	93.93%	93.93%	93.93%

Tabla 6.1: Resultados generales

Presentamos los resultados por clases sobre cada dataset en la Tabla 6.2 y realizamos un breve análisis. Los resultados de la técnica BKP para cada clase muestran que para estos datos la clase 6 es la que más le costó identificar. Los resultados de la técnica LBP para cada clase son muy alentadores ya que demuestran un excelente desempeño. Los resultados de la técnica HC para cada clase muestran que no pudo distinguirse muy acertadamente las clases 5 y 6. Los resultados de la técnica MC para cada clase permiten ver que para este conjunto de datos se puede identificar de forma ideal la clase 0.

Feature Extraction	Clase	A	P	R	FM
BKP	0	92.85%	80%	100%	88.88%
BKP	5	89.28%	100%	76.92%	86.95%
BKP	6	82.14%	62.5%	71.42%	66.66%
LBP	0	100%	100%	100%	100%
LBP	5	100%	100%	100%	100%
LBP	6	100%	100%	100%	100%
HC	0	92.85%	100%	81.81%	90%
HC	5	78.57%	66.66%	50%	57.14%
HC	6	78.57%	61.53%	88.88%	72.72%
MC	0	100%	100%	100%	100%
MC	5	92.85%	100%	81.81%	90%
MC	6	92.85%	81.81%	100%	90%

Tabla 6.2: Resultados de cada técnica de extracción de características para cada clase

6.2.4 Discusión y conclusiones

Algunas conclusiones que podemos extraer de los resultados son:

1. LBP es la técnica que presenta mejores resultados en este conjunto de datos ya que todas sus medidas son superiores al resto; lo que nos dice que la información local de textura tiene altas probabilidades de servir para caracterizar una imagen en base a su nivel de activación muscular
2. los segundos mejores resultados fueron obtenidos por la técnica MC, lo que nos indica que también la información de contorno sirve para esta caracterización
3. todos los resultados son superiores al 70% y muchos de ellos mayores al 90% lo que indica que las técnicas presentadas de extracción de características en combinación con la técnica de clasificación elegidas son útiles para estimar de forma discreta en 3 valores el nivel de carga a partir de imágenes estáticas
4. tenemos la intuición de que la causa del buen desempeño de LBP sea que la información de textura caracteriza mejor variaciones en la piel

En esta sección presentamos un enfoque que permite decidir con un alto grado de certeza qué nivel de carga muscular de entre 3 posibles está ejerciendo una persona a partir de una imagen de su brazo a piel descubierta. Esto fue posible realizando aprendizaje supervisado (SVM) sobre vectores de características visuales de las imágenes obtenidos aplicando diversas técnicas para tal fin (BKP, LBP, HC, MC). Para el conjunto de imágenes utilizadas en este trabajo, las técnicas de LBP y de MC dieron los mejores resultados. Sin embargo, todas las técnicas presentan una performance aceptable y superiores al 70% (llegando al 100% en el mejor caso). Cabe aclarar que el buen desempeño obtenido por todas las técnicas puede estar influenciado por la uniformidad en iluminación, escala, posición y ventana de recuadro impuesta a la captura de las imágenes de forma intencional con el objetivo de controlar lo mejor posible los datos de entrada. Para llevar este método a entornos más realistas, se debe buscar lograr generalización a estas dimensiones. Con ese objetivo es que en la próxima sección se presenta una aplicación práctica en tele-medicina donde se logra generalizar al punto de vista, usando información de nubes de puntos 3D como entrada.

6.3 Hacia una aplicación práctica en tele-medicina usando nubes de puntos 3D como entrada

6.3.1 El problema

En el Capítulo 5 se pretendía resolver un problema que requería un cierto nivel de generalización al pretender estimar activación para sujetos no vistos en entrenamiento. Los resultados en un setting simple no fueron de suficiente impacto y demostraron la necesidad de usar información específica del sujeto a evaluar. De esta manera en la primera sección de este capítulo se presentó un enfoque que dio muy buenos resultados en el problema de estimación intra-sujeto, pero con grandes limitaciones a nivel metodológico y de aplicación práctica. En particular, el setting de evaluación presentado anteriormente en la Sección 6.2 no propone ningún tipo de generalización ya que la evaluación se hace sobre imágenes muy similares a las usadas para entrenar, para mitigar esto en esta sección se pretende lograr generalización a ciertas variaciones en el punto de vista lo que aumenta la aplicabilidad práctica. Además, tanto en la sección anterior como en el Capítulo 5 se usa solo información 2D RGB del brazo, esta limitación se pretende superar en esta sección usando como entrada información 3D del brazo materializada en nubes de puntos 3D capturadas con dispositivos RGB-D económicos.

El contenido de esta sección ha sido publicado en [3] y el mismo aborda el problema de estimar remotamente la activación del músculo bíceps cuando se somete a fuerzas externas en contracción isométrica en condiciones de tele-fisioterapia: interiores, iluminación controlada e instalación simple. La instanciación del método presentado en esta sección, utiliza como entrada nubes de puntos 3D capturadas por dispositivos RGB-D. El sistema puede ser instalado por un paciente con pocos ejemplos de entrenamiento para calibrarlo para el mismo y producir estimaciones robustas para montajes imprecisos del hardware. Funciona suponiendo que la activación muscular expresa cambios en la geometría del brazo. Propone aprender el mapeo entre las características geométricas del brazo (caracterizado por el descriptor de nube de puntos *ESF* [99]) y el nivel de activación para el bíceps usando el algoritmo *SVM* [95, 15] con kernel linear. El método proporciona robustez a puntos de vista de medición no antes vistos causados por desvíos del aparato durante la configuración, o movimientos involuntarios del brazo durante la medición.

6.3.2 El enfoque

Esta sección presenta la metodología mencionada para la estimación no invasiva de la activación del músculo bíceps basada únicamente en información 3D geométrica de la deformación del brazo cuando se somete a un conjunto discreto de esfuerzos. Para producir los ejemplos de mapeo, se capturaron las nubes de puntos 3D del brazo (usando un dispositivo RGB-D Microsoft Kinect™) desde un punto de vista de medición base, mientras que la mano correspondiente sostenía objetos de pesos conocidos. Esos valores de peso se tomaron como medida del nivel de activación del bíceps. Se evaluó el desempeño del método caracterizando las nubes de punto con los descriptores globales de nubes de puntos presentados anteriormente en la Sección 4.3.2 para generar los vectores de características usados para aprender el modelo de mapeo.

El método descrito hasta ahora todavía presenta una deficiencia importante. En la práctica, el entrenamiento y posterior uso del sistema abarca días o semanas, por lo que es casi imposible garantizar una alineación precisa entre el punto de vista 3D del dispositivo mientras se producen los ejemplos de entrenamiento y el punto de vista cuando se capturan las nubes de puntos 3D para obtener el nivel de activación más adelante en el uso del sistema, por ejemplo en un proceso de rehabilitación. Esto da como resultado un sistema frágil, ya que los valores producidos por los descriptores geométricos son sensibles a los cambios en el punto de vista, porque se modifican qué partes de la superficie son visibles y cuáles están ocluidas. Para abordar esta grave deficiencia, el enfoque presentado en este capítulo propone ampliar la base de entrenamiento del algoritmo a más puntos de vista. El conjunto de ejemplos que se asignan a un peso dado se expande para incluir capturas del brazo también desde un conjunto de puntos de vista espacialmente vecinos, ubicados dentro de una esfera centrada en el punto de vista de medición base, con la expectativa de que el modelo aprendido ahora sea robusto para mediciones desde todos los puntos de vista dentro de esa esfera. En la práctica, esta propuesta es bastante exigente, si no completamente inviable, ya que requeriría que el paciente realice una enorme cantidad de levantamiento de objetos, y un tedioso re-posicionamiento del dispositivo de captura a todos los posibles puntos de vista del entrenamiento. Por lo tanto, el enfoque considera una generación simulada de ejemplos de entrenamiento que, basados solo en capturas desde un único punto

de vista, produce ejemplos para todos los puntos de vista vecinos. Para ello, en base a un pequeño conjunto de capturas reales, se generaron nubes de puntos como las que se observarían desde cada nuevo punto de vista, eliminando los puntos de la captura original que no deberían ser visibles desde este punto de vista, utilizando el operador *hidden point removal* [54]. Al calcular las características geométricas de las nubes de puntos para estos puntos de vista vecinos y unir el peso correspondiente de las capturas originales, se obtiene un ejemplo de entrenamiento simulado. En principio, las nuevas áreas visibles que ahora aparecen también deberían agregarse, pero aquí se ignoraron esas áreas adicionales. Como esto solo descarta información útil, agregarla de nuevo solo podría mejorar la calidad del modelo, por lo que los errores resultantes del modelo entrenado de esta manera deberían tomarse como un límite inferior.

El enfoque anteriormente presentado está representado gráficamente en la Figura 6.1.

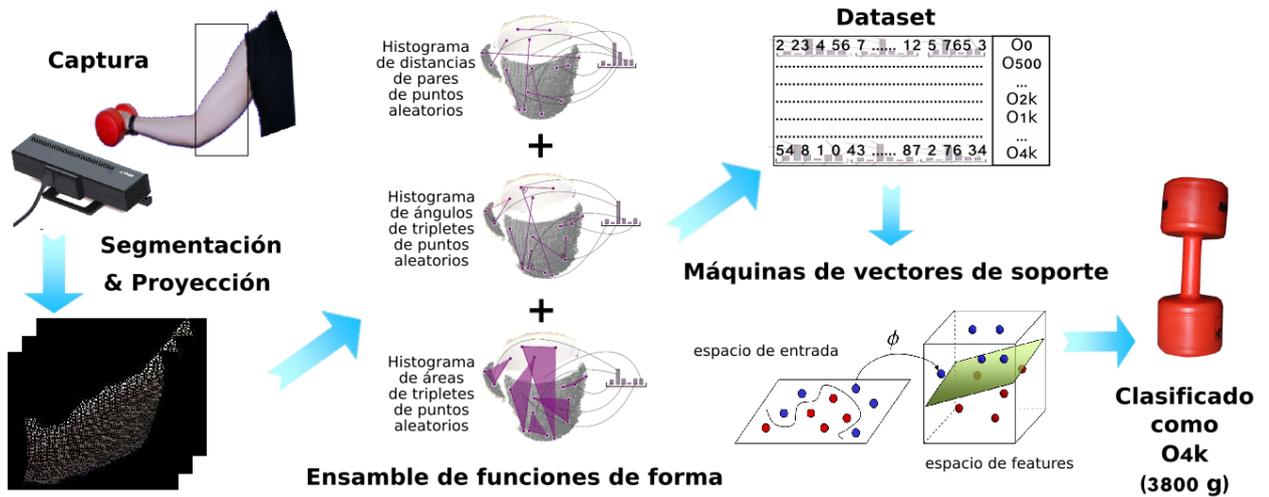


Figura 6.1: Enfoque presentado en esta sección

6.3.3 Evaluación empírica

Setup experimental

Esta sección describe los experimentos realizados para demostrar la efectividad del enfoque propuesto, medido en términos de su practicidad para resolver el problema de tele-fisioterapia, tomando como guía las dos aplicaciones descritas anteriormente en la Sección 2.3: (i) los ejercicios *mantener-relajar*, *contraer-relajar* y *mantener-relajar contracción agonista* que requieren controlar la sobre-contracción por encima de un porcentaje de MVC; y (ii) el *test muscular manual* que requiere la evaluación táctil de la contracción muscular. En ambos casos, la medición se basa en valores de activación discretos que oscilan entre ausencia de contracción y MVC. Por lo tanto, una estimación aproximada de activación es suficiente para mejorar una situación en la que los EMG son inviábiles. Según varios estudios [36, 11], la MVC de un músculo bíceps adulto está cerca de 350 N. Suponiendo un setting similar al de la Figura 6.2, para que el bíceps produzca tal fuerza a través de un peso en la mano bajo contracción isométrica, uno debe equiparar los pares producidos por cada fuerza [21]. Suponiendo que el peso en la mano W se aplica a aproximadamente 30 cm, la fuerza del músculo bíceps B a 4 cm y el peso del antebrazo G se supone que se aplica 15 cm desde la articulación del codo. Si el peso del antebrazo es de aproximadamente 15 N, produce un torque en la misma dirección que el peso de la mano, equivalente a 225 Ncm. Tal combinación da como resultado una fuerza de aproximadamente $W = (350 \times 4 - 225)/30 = 39.16$ N o 3.9 kg para producir 350 N de MVC en el bíceps. Por lo tanto, se consideraron cinco niveles discretos dentro de este rango de peso, aproximadamente duplicando en cada paso: 0 g, 505 g, 1340 g, 1940 g y 3885 g cada uno, nombrados aquí con las etiquetas O0, O500, O1k, O2k y O4k, respectivamente. Se puede suponer que un buen rendimiento del método que predice estos valores puede extrapolarse para esperar buenos resultados en casos prácticos. Sin embargo, se requerirían más pruebas antes de implementar este enfoque en tratamientos reales.

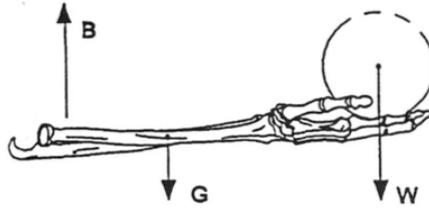


Figura 6.2: Setting de contracción isométrica considerado en esta tesis

El rendimiento del enfoque depende en gran medida de qué tan bien su conjunto de entrenamiento represente el mapeo subyacente entre las características geométricas del brazo y su nivel de activación. Para evaluar este rendimiento, se informa el *MAE* en gramos para un rango representativo de posibles escenarios de entrenamiento que dependen de varias alternativas elegidas por el usuario. Por un lado, la elección del descriptor de nube de puntos utilizado para generar vectores de características: VFH, CVFH o ESF. Por otro lado, las elecciones relativas al aumento basado en simulación del conjunto de datos: el número n de las capturas originales; el *trainRadius* de la esfera; y el número N de puntos de vista vecinos simulados.

Utilizando un dispositivo Microsoft Kinect™, se capturaron nubes de puntos 3D desde una vista frontal del brazo (ver Figura 6.3) colocándolo en la coordenada $(0, 0, -0.2)$, que corresponde a 20 cm sobre el eje Z negativo y apuntando hacia el origen. Todas las capturas se mantuvieron con iluminación y escala uniforme en una postura de contracción isométrica de bíceps. El sujeto mantuvo el segmento del brazo del hombro y el codo aproximadamente 45° con respecto al eje X y el segmento del brazo del codo paralelo al eje X. Todo el brazo estaba apoyado solo en el hombro por el respaldo de una silla sobre la que estaba sentado el sujeto.

Para cada uno de los cinco pesos, se capturaron cuatro ráfagas usando una herramienta de la biblioteca Point Cloud Library [83] a 30 fps. Se capturaron diez segundos para un total de 300 cuadros, de los cuales se tomaron 25 al azar, lo que resultó en cuatro grupos de 25 nubes de puntos 3D por peso. Se usaron tres ráfagas por peso ($75 \times 5 = 375$ nubes de puntos) para construir el conjunto de entrenamiento. Una ráfaga por peso ($25 \times 5 = 125$ nubes de puntos) se separó para construir el conjunto de pruebas siguiendo la típica división de train/test 75%/25%.

A continuación, estas nubes de puntos se procesaron manualmente segmentando los puntos que no corresponden a los bíceps y moviendo la nube de puntos resultante, de modo que el punto en los bíceps correspondiente a la distancia media entre el codo y el hombro se ubicó en el origen de el sistema de coordenadas. Se eligieron los mismos planos de segmentación para todas las imágenes de la misma ráfaga. Para un sistema final, la segmentación podría automatizarse con algoritmos de detección de objetos tridimensionales [72, 96] entrenados a través de ejemplos para ubicar los bíceps en la nube de puntos completa de la escena.

Las nubes de puntos resultado del proceso anteriormente explicado son tomadas como inicio del cual se parte para generar el resto de las nubes de entrada a través de procesos de simulación.

Resultados experimentales

Selección del descriptor geométrico

El mejor descriptor sería aquel con el mayor poder discriminatorio, solo medible al comparar su desempeño con algún algoritmo de aprendizaje supervisado. Para hacer esto, el MAE obtenido por SVM con kernel lineal se evaluó sobre 75 puntos de vista base de medición diferentes. Estos puntos se consideraron distribuidos uniformemente en una grilla de 40 cm por 40 cm por 20 cm dentro del prisma encerrado por los vértices ($X = 20$ cm, $Y = 20$ cm, $Z = -20$ cm) hasta el punto 3D ($X = -20$ cm, $Y = -20$ cm, $Z = 0$ cm), distribuido uniformemente cada 10 cm en las tres direcciones. Se produjo un par de conjunto de entrenamiento y de testeo por punto de vista para cada uno de los tres descriptores geométricos. Los experimentos también se realizaron con un four-fold cross validation sobre el hiperparámetro C , considerando los siguientes valores posibles: $C = \{2 \times 10^{-9}, 2 \times 10^{-7}, 2 \times 10^{-5}, 2 \times 10^{-3}, 2 \times 10^{-1}, 2, 2 \times 10^1, 2 \times 10^3, 2 \times 10^5, 2 \times 10^7, 2 \times 10^9\}$.

La Figura 6.4 muestra la diferencia en MAE de los descriptores VFH y CVFH con ESF para cada uno de los 75 puntos de vista, donde las barras de más de 0 g muestran que ESF se desempeñó mejor para el punto

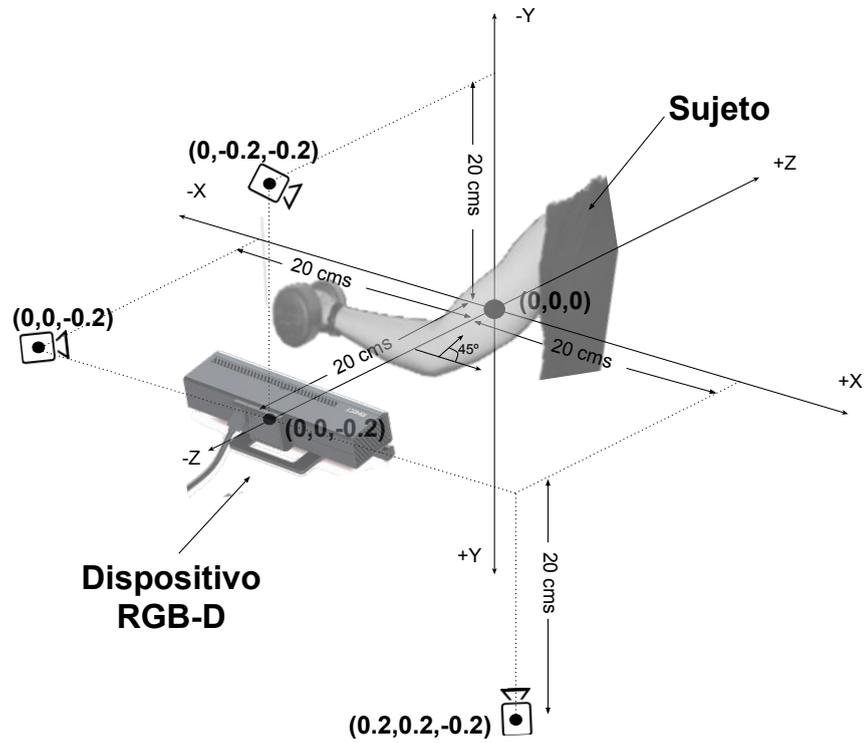


Figura 6.3: Setup de captura

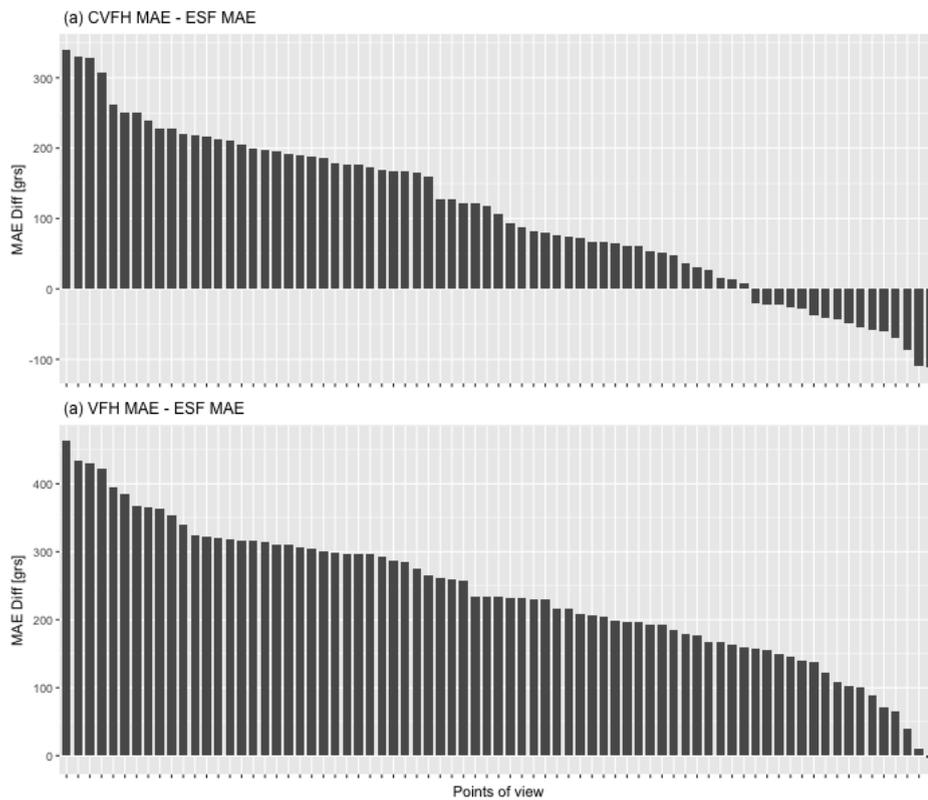


Figura 6.4: (a) muestra la diferencia del MAE de CVFH menos el de ESF para cada uno de los puntos de vista evaluados; (b) muestra la diferencia en MAE de VFH menos ESF para cada uno de los puntos de vista evaluados.

de vista correspondiente. La Figura 6.4 (a) muestra que ESF se desempeñó mejor que CVFH en la mayoría de los puntos de vista evaluados. La Figura 6.4 (b) muestra que ESF obtuvo un MAE mejor que VFH en casi todos los puntos de vista probados. En base a estos resultados, se utilizó ESF como el descriptor geométrico de elección para los experimentos restantes.

Eligiendo los parámetros óptimos para la expansión de los datos

En los siguientes experimentos, se consideraron escenarios de aprendizaje sobre cuatro posibles *puntos de vista de medición*, todos de ellos apuntando al origen de coordenadas como muestra la Figura 6.3:

- $(0, 0, -0.2)$: 20 cm frente al bíceps ,
- $(0, -0.2, -0.2)$: 20 cm frente y 20 cm sobre el bíceps,
- $(-0.2, 0, -0.2)$: 20 cm frente y 20 cm a la izquierda del bíceps,
- $(0.2, 0.2, -0.2)$: 20 cm frente, 20 cm a la derecha , y 20 cm bajo el bíceps.

Para producir los conjuntos de entrenamiento para estos escenarios de aprendizaje, se realizó el siguiente proceso:

1. Considerando las $n = \{1, 2, 3\}$ ráfagas de entrenamiento capturadas, se simularon n nubes de puntos 3D base para cada uno de los puntos de vista de medición
2. El conjunto de entrenamiento se construyó al aumentar el conjunto base a través de la simulación de nubes de puntos 3D nuevas para los puntos de vista en el entorno espacial del *punto de vista de medición*, para lo que se muestrearon uniformemente N puntos de vista dentro de una esfera de radio $trainRadius$ centrado en el *punto de vista de medición*
3. Para cada uno de estos N puntos de vista vecinos y cada ráfaga de entrenamiento n , se tomaron muestras de dos nubes de puntos a partir de las nubes de puntos base, desde las cuales se generaron nuevas nubes de puntos simuladas para ese punto de vista, lo que resultó en $2 \times n \times N$ nubes de puntos simuladas por punto de vista de medición
4. Teniendo en cuenta todas estas nubes de puntos, el conjunto de entrenamiento se generó calculando el vector de características geométricas, usando *ESF*, para cada una y anexándolas el peso levantado como etiqueta

Considerando el proceso anterior, los siguientes experimentos informan resultados para diferentes valores de n , N y $trainRadius$.

Los conjuntos de prueba también se generaron mediante simulación utilizando al azar dos nubes de la cuarta ráfaga reservadas para fines de prueba. En este caso se tomaron 100 puntos de vista de una *circunferencia* de radio $testRadius$ centrada en el punto de vista de medición cuyo plano es perpendicular a la línea de visión (la línea que conecta la ubicación del punto de vista de medición con el origen del sistema de coordenadas). Esta opción de escenario de prueba representa los errores esperados en la colocación del Kinect, para los cuales la distancia es mucho más fácil de garantizar que el ángulo. A partir de todas estas nubes de puntos simuladas, el conjunto de prueba se generó calculando el vector de características geométricas para cada una y anexándolas el peso levantado como etiqueta.

Convergencia en n y N

El número N de puntos de vista simulados y las n ráfagas, influyen en el tamaño del conjunto de entrenamiento, y debe demostrarse convergencia para concluir que se han producido suficientes data-points de entrenamiento. Para n , se espera un bajo valor de convergencia, ya que esas ráfagas deben capturarse manualmente para cada uno de los pacientes. En contraste, las capturas simuladas N se obtienen computacionalmente, por lo que no hay impacto en la practicidad del método para valores altos de N .

La Figura 6.5 demuestra empíricamente que $n = 3$ ráfagas son suficientes para la convergencia cuando el conjunto de entrenamiento se aumenta con $N = 1000$ puntos de vista simulados en una esfera de $trainRadius = 2$ cm. La figura muestra un gráfico de barras por punto de medición, cada uno mostrando el MAE para los posibles valores $n = 1, 2, 3$ (representados en barras de tonos crecientes de gris) sobre los

escenarios de prueba para valores crecientes de *testRadius*: 3 mm, 2 cm, 4 cm, 7 cm y 12 cm. Como se observa en la Figura 6.5, en todos los casos hay una marcada disminución en MAE entre una y dos ráfagas, mientras que ninguna mejora importante o incluso una disminución en la calidad (principalmente para valores grandes de *testRadius*) ocurre entre dos y tres ráfagas. Esta convergencia es una señal clara de que más de 3 ráfagas no mostrarían una mejora significativa.

Para decidir cuántos puntos de vista simulados N son suficientes, el *trainRadius* también se fijó en 2 cm y se informó el MAE para un número creciente de N dentro de la esfera correspondiente, $N = \{20, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000\}$ para los cuatro puntos de vista de medición. Para mejorar aún más la representatividad, se generaron diez conjuntos de entrenamiento por N . Para cada conjunto de entrenamiento, se aprendió y probó un mapeo para escenarios de diferentes *testRadius*: 3 mm, 2 cm, 4 cm, 7 cm y 12 cm. La Figura 6.6 muestra los resultados de cada punto de vista de medición que muestra una curva por *testRadius*. Cada figura traza la media del MAE y la desviación estándar para cada uno de los cinco escenarios de prueba.

Esto muestra que más data-points disminuyen el MAE, al menos para N pequeños, junto con la convergencia de $N = 500$ para los cuales, a pesar de algunas oscilaciones menores, el promedio de MAE permanece constante. Esto corresponde al hecho de que a aproximadamente $N = 500$, no se puede esperar una mejora adicional en la calidad del mapeo.

Impacto del *trainRadius* en la performance del método

Los experimentos realizados fijaron hasta ahora el *trainRadius* a 2 cm. Esta sección muestra las tendencias al aumentar los valores de este radio, mediante la evaluación de cómo los modelos aprendidos de los conjuntos de entrenamiento expandidos espacialmente tienen un impacto en la medición con potenciales desalineamientos en los dispositivos de medición. Por lo tanto, se generaron conjuntos de entrenamiento para *trainRadius* = {0, 0.5, 1.0, 1.5, 2.0} cm. Un modelo para cada uno fue aprendido y probado para escenarios de diferentes *testRadius*: 3 mm, 2 cm, 4 cm, 7 cm y 12 cm. Por coherencia, se mantuvo la misma densidad de puntos por esfera de entrenamiento en lugar de la misma cantidad de puntos (lo que resultó en un aumento cúbico en los N reales). $N = 1000$ puntos de vista se simularon para el valor más grande de *trainRadius* = 2 cm, y se submuestrearon nubes para las esferas más pequeñas. La Figura 6.7 muestra el MAE para los cuatro puntos de vista de medición, con cada figura que contiene una curva por *testRadius* que abarca todos los valores de *trainRadius*.

Para todos los puntos de vista de medición y todos los *testRadius* en cada figura, las curvas muestran una clara tendencia a reducir el MAE con un aumento de *trainRadius*, lo que demuestra empíricamente que la expansión espacial del conjunto de entrenamiento produce un mejor mapeo, como lo muestran los valores más bajos de MAE. Además, solo la cifra correspondiente a $(-0.2, 0, -0.2)$ muestra convergencia, lo que demuestra que en todos los demás casos puede haber más mejoras para expansiones espaciales incluso mayores del conjunto de entrenamiento.

6.3.4 Discusión y conclusiones

Los resultados experimentales presentados prueban la existencia de correlaciones entre la activación muscular y las imágenes tridimensionales del bíceps susceptibles de ser extraídas por sistemas autónomos no intrusivos y lo suficientemente simples para ser instalados fácilmente por técnicos no médicos para ser operados posteriormente por el usuario.

Los resultados de los experimentos muestran que este enfoque es capaz de encontrar patrones discriminatorios entre niveles discretos de activación, que van desde cero hasta aproximadamente MVC, con errores lo suficientemente bajos como para competir con un EMG. Para ampliar este argumento la Tabla 6.3 muestra los resultados anteriores para *trainRadius* = 2 cm y $n = 3$ ráfagas de entrenamiento (los valores para estos parámetros que mostraron los mejores resultados en todos los casos) todos los puntos de vista de medición, y todos los *testRadius*. Por cada *testRadius*, la tabla muestra el valores de la medida de de rendimiento en el valor de N con el mejor MAE (N^*). La Tabla 6.3 (b) muestra los mejores resultados para el punto de vista $(0, -0.2, -0.2)$ correspondiente a una vista frontal superior del bíceps. Para este punto se muestra que las mediciones de hasta 4 cm desde el punto de vista de medición incurrirían en un MAE menor a 350 g y las mediciones de hasta 12 cm desde el punto de vista de medición incurrirían en un MAE menor a 450 g . Teniendo en cuenta que en el problema bajo estudio, el MVC corresponde al mayor peso de 3885 g, los errores obtenidos con este método están entre el 9 % (350 g) y el 11 % (450 g) de MVC. Aunque

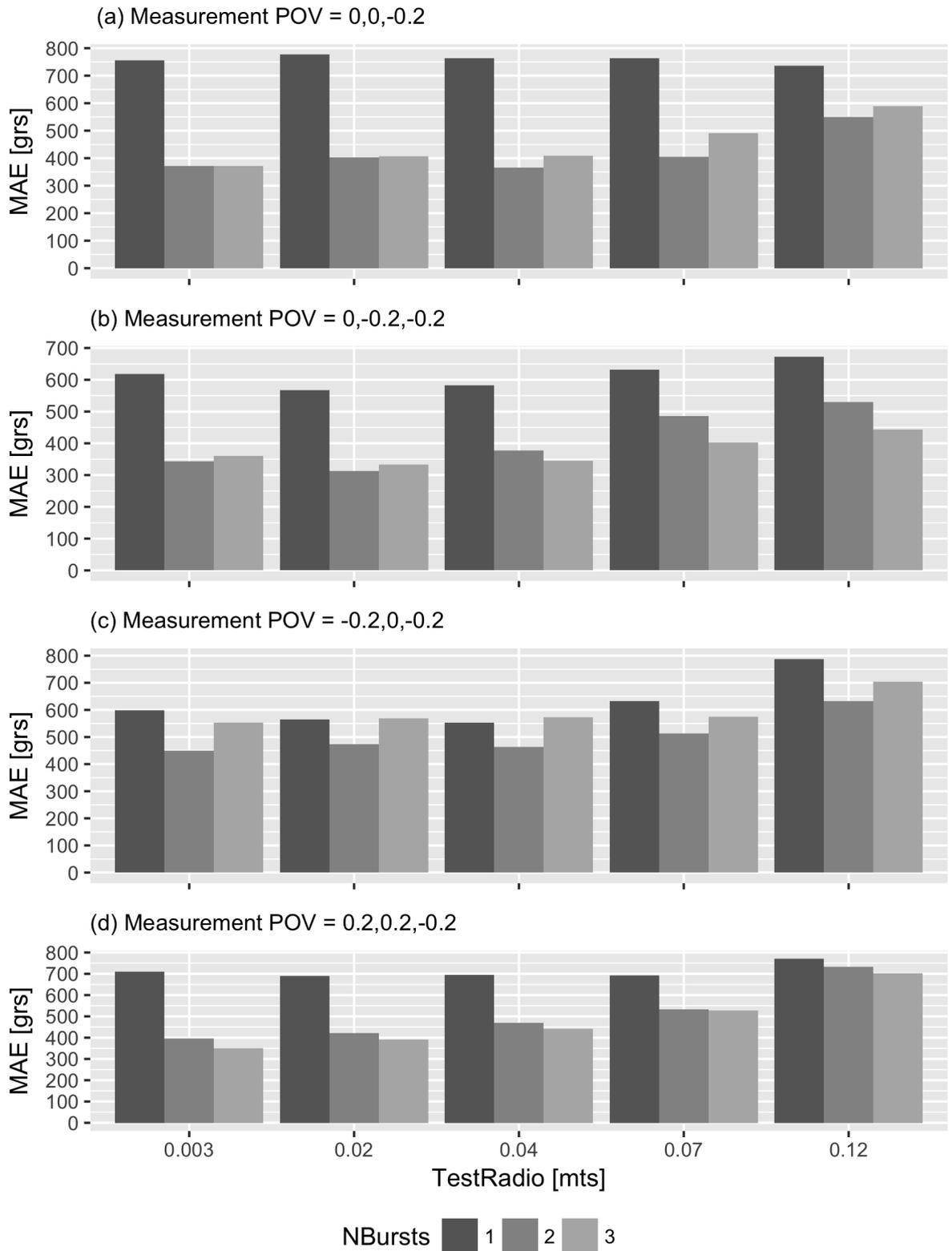


Figura 6.5: Las Figuras (a), (b), (c), y (d) muestran el MAE para valores crecientes de n ráfagas de entrenamiento para un $trainRadius = 2$ cm y $N = 1000$ sobre los cuatro puntos de vista de medición.

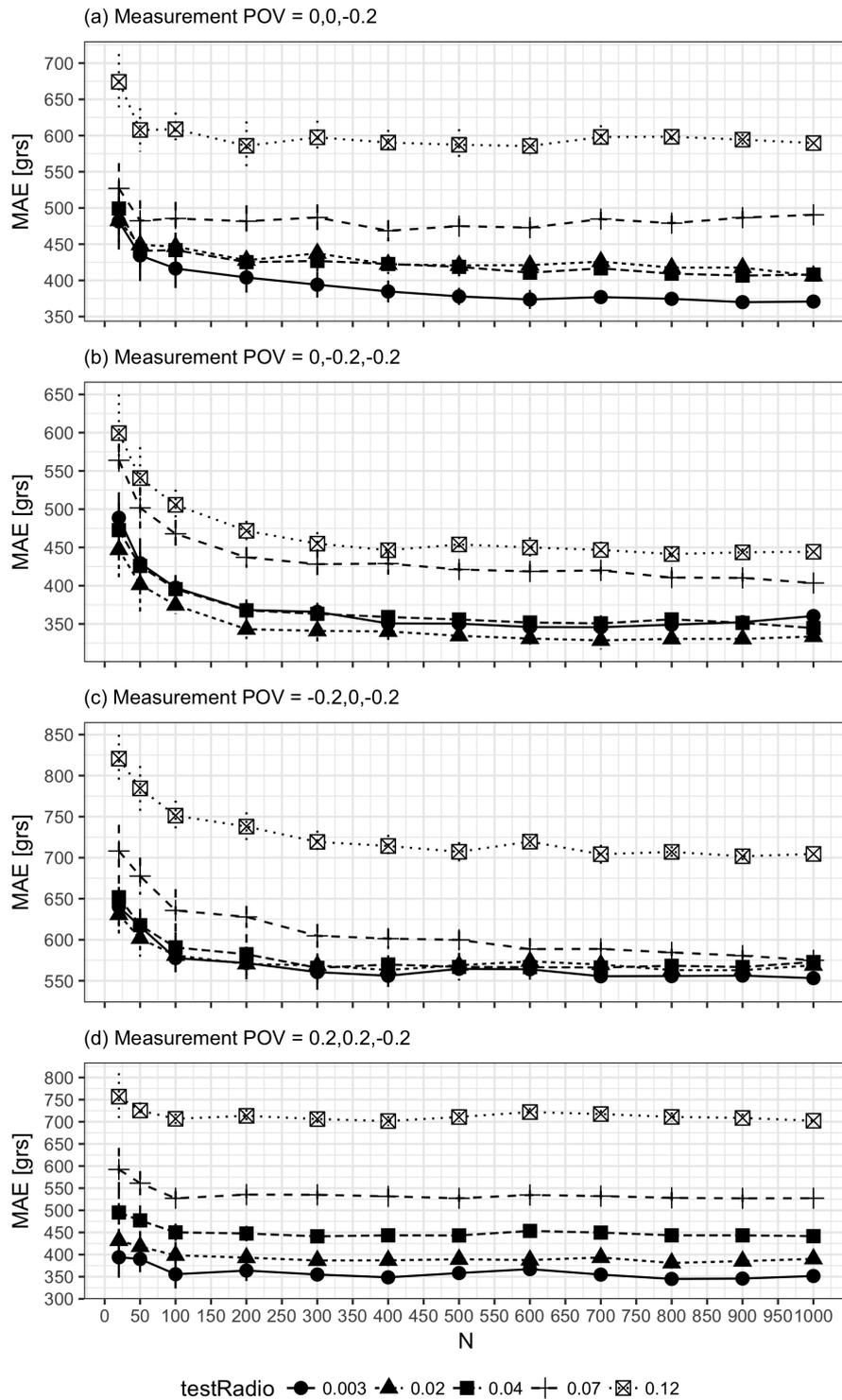


Figura 6.6: Las Figuras (a), (b), (c) y (d) muestran el promedio y desviación estándar del MAE para un $trainRadius = 2\text{cm}$, valores crecientes de N sobre los cuatro puntos de vista de medición, con una curva para cada uno de los cinco $testRadius$.

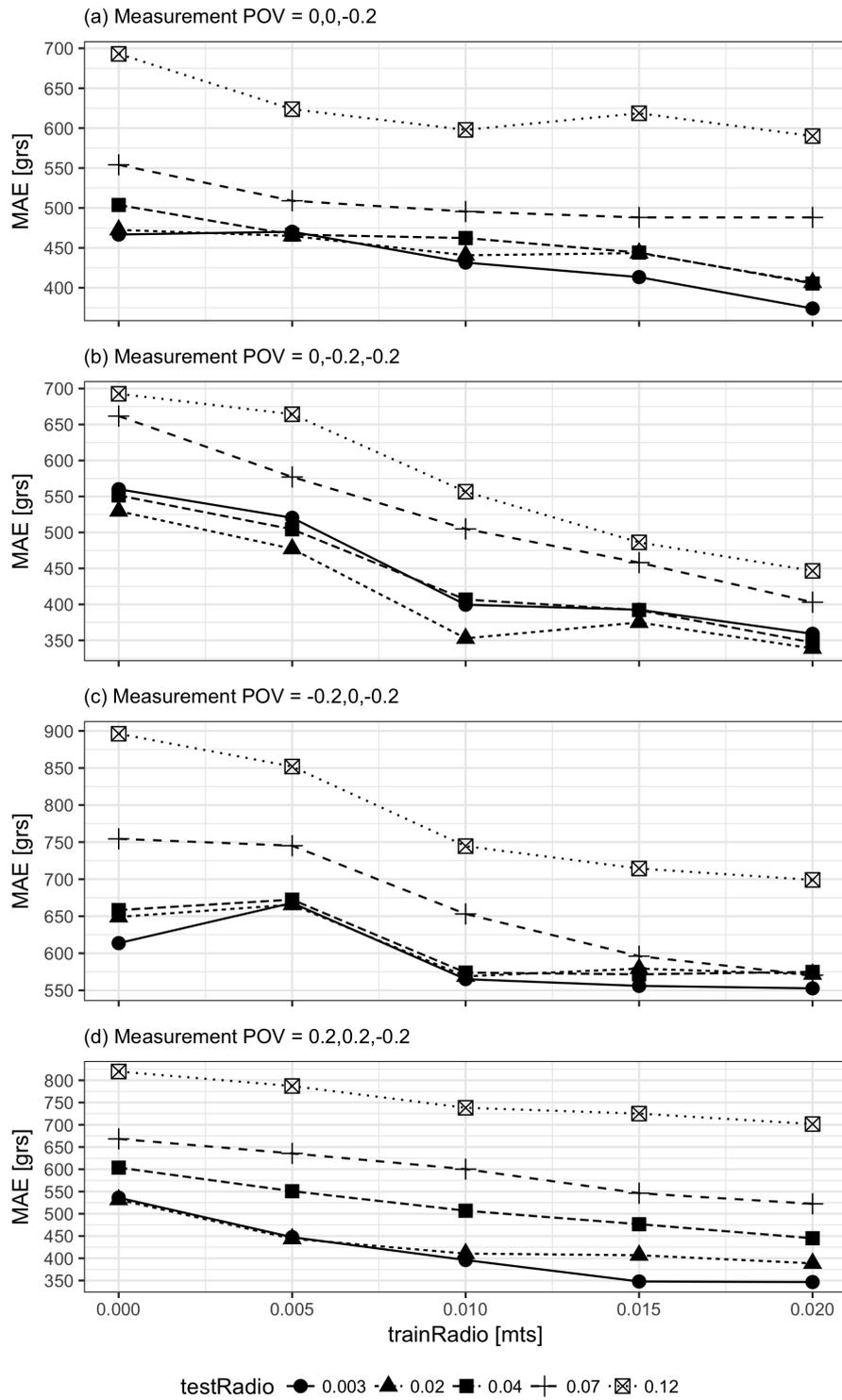


Figura 6.7: Las Figuras (a), (b), (c), y (d) muestran el promedio del MAE para valores incrementales de *trainRadius* sobre los cuatro puntos de vista de medición, con una curva por *testRadius*.

no se encontraron trabajos iguales que estiman el nivel de activación exclusivamente a partir de nubes de puntos 3D, los errores obtenidos por el método propuesto son equivalentes a los informados para mediciones de fuerza con EMG [91, 78] que oscilan entre 9 % y 10 % del valor MVC.

Como se presentó en el Capítulo 3, similar a este trabajo, hay un grupo de trabajos que abordan el problema de predecir la activación muscular, pero a partir de información cinemática y cinética. Analizaremos el desempeño de algunos de los trabajos más recientes de este grupo para compararlos con los resultados obtenidos por nuestro método. Los mejores resultados de [87] muestran un coeficiente de determinación de 0.42 entre la activación estimada y real. En [94] los mejores resultados presentan un coeficiente de determinación de 0.6, y específicamente para el bíceps 0.5. En [101] el coeficiente de determinación de todas las ecuaciones de regresión oscila entre 0.228 y 0.818. En [67] los mejores resultados muestran un coeficiente de determinación superior a 0.9.

Para los trabajos revisados, la información de posiciones y ángulos articulares es altamente informativa, sin embargo, podría ser el caso de que para la misma postura o movimiento, el nivel de contracción puede ser diferente, y usar solo la cinemática para inferir la activación puede fallar. Por esta razón, agregan información de fuerza externa como entrada a sus modelos. Aunque la medición de la fuerza es útil, en un entorno de tele-fisioterapia, la adición de sensores de fuerza es un requisito poco práctico que cumplir. Por el contrario, el enfoque propuesto mide el efecto de las fuerzas externas sobre la activación de forma indirecta y no invasiva al considerar la deformación de la superficie del brazo. Además, dado que se estudian posturas estáticas, la información cinemática en este enfoque no es necesaria ya que es un valor constante. Sin embargo, si se analizara el movimiento, definitivamente se debe incluir. Aunque el problema abordado en este trabajo es uno de clasificación y estos trabajos relacionados estiman el nivel continuo de activación muscular, para comparar de algún modo los resultados con ellos, el coeficiente de determinación R^2 también se calculó transformando las clases en sus valores en gramos e informados en la Tabla 6.3. Los trabajos presentados informan coeficientes de determinación entre 0.228 y 0.9 y exclusivamente para los bíceps, el valor obtenido por [94] fue de 0.5. El coeficiente de determinación informado para el mejor caso de este trabajo es de entre 0.58 y 0.7, lo que muestra que los resultados de este método son comparables con los informados por los trabajos revisados. Estos resultados son alentadores teniendo en cuenta que, aunque se estiman niveles discretos para posturas estáticas, la tarea que se enfrenta es más desafiante, porque la cinemática junto con las fuerzas externas, definen casi por completo la activación muscular y son más informativas que la información geométrica solamente.

A partir de este análisis, se puede concluir que el trabajo propuesto en esta sección constituye un resultado positivo hacia una medición práctica del nivel discreto de activación del músculo bíceps en el contexto de tele-rehabilitación.

testRadius	N^*	MAE	R^2
0.3	900	369.98 (7.31)	0.63 (0.01)
2	1000	406.20 (0)	0.6 (0)
4	900	406.62 (4.57)	0.6 (0)
7	400	468.55 (14.51)	0.52 (0.02)
12	600	585.53 (12.10)	0.39 (0.01)

(a) Measurement POV = 0,0,-0.2

testRadius	N^*	MAE	R^2
0.3	1000	553.20 (0)	0.43 (0)
2	900	562.80 (3.5)	0.42 (0)
4	300	565.78 (17.5)	0.45 (0.02)
7	1000	574.95 (0.6)	0.43 (0)
12	900	701.76 (8.21)	0.25 (0.01)

(c) Measurement POV = -0.2,0,-0.2

testRadius	N^*	MAE	R^2
0.3	700	345.73 (15.42)	0.67 (0.01)
2	700	328.55 (11.73)	0.7 (0.01)
4	1000	344.60 (0)	0.68 (0)
7	1000	403.50 (0)	0.64 (0)
12	800	441.43 (9.34)	0.58 (0.01)

(b) Measurement POV = 0,-0.2,-0.2

testRadius	N^*	MAE	R^2
0.3	800	344.98 (7.71)	0.6 (0.01)
2	800	381.12 (12.34)	0.56 (0.01)
4	300	441.20 (12.99)	0.5 (0.02)
7	900	527.02 (8.32)	0.41 (0.01)
12	400	701.68 (11)	0.13 (0.03)

(d) Measurement POV = 0.2,0.2,-0.2

Tabla 6.3: MAE y R^2 del método para $n = 3$ ráfagas de entrenamiento y $trainRadius = 2cm$ para el mejor N para cada $testRadius$ (reportado en cms). Las Tablas (a), (b), (c) y (d) para cada uno de los puntos de vista de medición.

Esta sección presenta los primeros pasos en la tarea de estimación del nivel de activación del músculo bíceps a partir de nubes de puntos 3D del brazo utilizando visión computacional y aprendizaje de máquinas supervisado. Su objetivo principal es la exploración temprana de estas tecnologías para resolver el problema que satisface los requisitos de un entorno de tele-medicina. La solución al problema específico abordado en

este capítulo para medir niveles discretos de activación de bíceps en contracción isométrica podría ponerse en práctica, con pequeñas mejoras, en la tarea de control de intensidad y tiempo de contracción muscular en ejercicios de tele-rehabilitación de contracción isométrica de bíceps y como una forma para automatizar la ejecución del *test muscular manual* para medir la salud del bíceps. El enfoque consiste en un método que considera la generación de un vector de características para las nubes de puntos del brazo utilizando el descriptor geométrico 3D *conjunto de funciones de forma* (ESF) para una imagen 3D dada capturada en un punto de vista recomendado y su vecindad espacial. A continuación, utiliza un modelo entrenado utilizando la técnica de aprendizaje automático supervisada *máquinas de vectores de soporte* para estimar de forma autónoma el nivel de activación de nuevos ejemplos. Un hallazgo importante es que aumentar espacialmente la cantidad y variabilidad de los datos de entrenamiento ayuda a aumentar la precisión de la estimación y a reducir el número de capturas reales requeridas para entrenar un modelo con suficiente precisión, y brinda robustez a errores en el punto de vista de medición producidos por un montaje inadecuado. Aunque este es un primer paso sólido para lograr una aplicación práctica, la eficiencia obtenida se limita a una contracción isométrica y al músculo bíceps.

6.4 Conclusiones

En este capítulo se presentaron resultados experimentales de la evaluación del método propuesto en la tarea de aprender modelos personalizados por sujeto y estimar la activación con estos modelos. Puntualmente la instancia del método presentada en la Sección 6.3 demuestra la factibilidad del uso de las tecnologías evaluadas en una aplicación práctica relativamente inmediata en el campo de la tele-rehabilitación.

En el próximo capítulo se cierra esta tesis presentando conclusiones generales del trabajo.

Capítulo 7

Conclusiones y trabajo futuro

En esta tesis se propone atacar el problema de estimar el nivel de activación discreto que ejerce el músculo bíceps de un brazo en postura de contracción isométrica, usando como información de entrada solamente imágenes externas del brazo. El enfoque propuesto para resolver el problema es atacarlo como uno de clasificación de imágenes. El mismo consiste en aplicar técnicas de visión computacional para caracterizar las imágenes de entrada en vectores de características. Estos vectores son proveídos como entrada a algoritmos de aprendizaje de máquinas supervisados con el nivel de activación discreto ejercido en cada imagen de entrada de manera de poder aprender un modelo que haga un mapeo de la imagen con el nivel de activación. Como etiqueta para el aprendizaje se utilizan objetos de pesos conocidos que permiten asegurar que el bíceps ejerza niveles discretos de activación entre un esfuerzo mínimo necesario para sostener el antebrazo y la contracción máxima voluntaria. En esta tesis se evaluó el método general en dos variantes del problema: (a) estimar el nivel de activación de un sujeto usando un modelo construido con imágenes de otros sujetos y (b) estimar el nivel de activación usando modelos personalizados aprendidos para cada sujeto en particular.

En la primera evaluación del enfoque se utilizan como entrada imágenes RGB 2D, las cuales se caracterizan con los descriptores RGB *patrones binarios Locales e histogramas de color HS*. En esta primera variación se evaluó la performance de los algoritmos *máquinas de vectores de soporte y bosques aleatorios*. Los resultados de la experimentación realizada demuestran la complejidad del problema a partir de los cuales se pueden destacar las siguientes principales conclusiones:

- la generalización a sujetos no vistos en el entrenamiento del sistema no se alcanza con niveles suficientes de precisión debido a las particularidades visuales de cada brazo
- mediante un análisis de curvas de aprendizaje, se estima que se requieren una cantidad enorme de sujetos de entrenamiento (al menos muchos mas de 90) para lograr la generalización deseada
- es posible disminuir esta complejidad muestral (la cantidad de sujetos de entrenamientos necesarios), si se incluye información previa del sujeto sobre el que se va a evaluar el modelo, que en el caso puntual de esta tesis fueron imágenes correspondientes al nivel de esfuerzo mínimo para el sujeto en evaluación como parte de la entrada a los modelos
- si bien esta *calibración* mejora la performance de esta variación método, aún la misma no es aceptable lo que da indicios de que siguen siendo necesarios más datos de entrenamiento
- es posible que la información de textura y color extraída de las imágenes RGB 2D no sea suficiente para caracterizar el estado de deformación real 3D, al menos no con imágenes estáticas

Para validar si es posible mejorar los resultados, se propone evaluar el método en un problema un poco más simple pero con impacto práctico válido. Estimar el nivel de activación muscular que ejerce un sujeto usando modelos específicos para ese sujeto. En una primera instancia se evaluó el desempeño del uso de imágenes RGB en esta tarea, considerando además de los mencionados *patrones binarios locales e histogramas de color HS* para describir las imágenes RGB, el uso de los descriptores *bolsa de puntos clave y momentos de contorno*. A pesar de que los resultados obtenidos fueron excelentes, esto se debe a la uniformidad de los datos lo que no propone ningún desafío de generalización. Para superar algunos de los inconvenientes presentados por la primera instancia evaluada del enfoque es que se propone otra instanciación del método.

La primera variación es que utiliza como entrada nubes de puntos 3D, caracterizadas por el descriptor global de nubes de puntos *conjunto de funciones de forma*. Para esta variación se usó como algoritmo de aprendizaje *maquinas de vectores de soporte*. Además, esta instanciación del método también aprende un modelo personalizado para el sujeto a evaluar, pero pretendiendo generalizar a variaciones en el puntos de vista. Debido a que esta variación se pensó para una aplicación práctica en tele-medicina es que se necesita ese nivel de generalización para darle robustez a posibles diferencias entre el punto de vista de testeo (cuando el sistema esté en uso) respecto del punto de vista central de entrenamiento (usado para calibrar el sistema). Esta robustez se logró ampliando la base de entrenamiento con ejemplos simulados de capturas desde puntos de vistas espacialmente vecinos del punto de vista central captura. Los resultados obtenidos por esta variante del método son aceptables para entornos de tele-medicina donde no se necesita una precisión demasiado alta y donde es posible la estimación en niveles discretos de activación muscular.

Los resultados y métodos presentados en esta tesis son los primeros pasos en la literatura en la tarea de resolver el problema usando las tecnologías mencionadas y permiten dimensionar la complejidad del mismo y la viabilidad del uso de estas tecnologías para resolverlo. El método y la experimentación realizada presentan aún limitaciones, principalmente por la restricción a posturas estáticas de contracción isométrica. A pesar de eso, dada la complejidad del problema de intentar inferir el nivel de activación muscular solo a partir de lo que se puede observar visualmente (tarea desafiante incluso para un humano) los resultados son aceptables aunque demuestran que para llevar este sistema a situaciones más realistas, probablemente se necesite otro tipo de información así como técnicas de visión computacional y posiblemente de aprendizaje de máquinas, más complejas. A pesar de esto, consideramos que la hipótesis planteada en esta tesis se ha validado ya que los experimentos demuestran una buena correlación entre lo externamente observable del brazo, y su nivel de activación muscular interno.

Como trabajo futuro se pretende encontrar otras situaciones donde el método presentado puede ser de utilidad práctica. Para esto estamos empezando a colaborar con empresas dedicadas al desarrollo de productos y servicios relacionados a la tele-medicina y tele-rehabilitación de manera de lograr que ellos puedan aplicar lo aprendido en esta tesis al desarrollo de productos innovadores, así como a otras situaciones y ejercicios además de los presentados en esta tesis. Al intentar esta transferencia puede ser posible que el método deba ser mejorado para contemplar el análisis de movimiento. Además resta por experimentar la aplicación de tecnologías más novedosas en el área de visión computacional y aprendizaje de máquinas como son los enfoques de aprendizaje profundo, los cuales han demostrado resolver el problema conjunto de descripción y aprendizaje de modelo de forma conjunta, siempre contando con grandes volúmenes de datos.

Referencias

- [1] Estimación de carga muscular mediante imágenes. Sociedad Argentina de Informática, Buenos Aires, Argentina (2014), <http://43jaiio.sadio.org.ar/proceedings/ASAI/search.html>
- [2] Abraham, L., Bromberg, F., Forradellas, R.: Arm muscular effort estimation from images using computer vision and machine learning. In: *Ambient Intelligence for Health*. pp. 125–137. Springer (2015)
- [3] Abraham, L., Bromberg, F., Forradellas, R.: Ensemble of shape functions and support vector machines for the estimation of discrete arm muscle activation from external biceps 3d point clouds. *Computers in Biology and Medicine* 95, 129 – 139 (2018), <http://www.sciencedirect.com/science/article/pii/S0010482518300416>
- [4] Aldoma, A., Vincze, M., Blodow, N., Gossow, D., Gedikli, S., Rusu, R.B., Bradski, G.: Cad-model recognition and 6dof pose estimation using 3d cues. In: *Computer Vision Workshops (ICCV Workshops)*, 2011 IEEE International Conference on. pp. 585–592. IEEE (2011)
- [5] Amarantini, D., Rao, G., Berton, E.: A two-step emg-and-optimization process to estimate muscle force during dynamic movement. *Journal of Biomechanics* 43(9), 1827–1830 (2010)
- [6] Amari, S.I., Murata, N.: Statistical theory of learning curves under entropic loss criterion. *Neural Computation* 5(1), 140–153 (1993)
- [7] Andersen, L.L., Andersen, C.H., Mortensen, O.S., Poulsen, O.M., Bjørnlund, I.B.T., Zebis, M.K.: Muscle activation and perceived loading during rehabilitation exercises: comparison of dumbbells and elastic resistance. *Physical therapy* 90(4), 538–549 (2010)
- [8] Anderson, C.V., Fuglevand, A.J.: Probability-based prediction of activity in multiple arm muscles: implications for functional electrical stimulation. *Journal of neurophysiology* 100(1), 482–494 (2008)
- [9] Assassi, L., Becker, M., Magnenat-Thalmann, N.: Dynamic skin deformation based on biomechanical modeling. In: *Proc. of the 25th CASA*. vol. 2 (2012)
- [10] Basmajian, J.V., De Luca, C.: *Muscles alive*. Baltimore/USA (1985)
- [11] Baudry, S., Rudroff, T., Pierpoint, L.A., Enoka, R.M.: Load type influences motor unit recruitment in biceps brachii during a sustained contraction. *Journal of neurophysiology* 102(3), 1725–1735 (2009)
- [12] Bauer, A., Paclet, F., Cahouet, V., Dicko, A.H., Palombi, O., Faure, F., Troccaz, J., et al.: Interactive visualization of muscle activity during limb movements: Towards enhanced anatomy learning. In: *Eurographics Workshop on VCBM* (2014)
- [13] Bay, H., Tuytelaars, T., Van Gool, L.: Surf: Speeded up robust features. In: *Computer Vision–ECCV 2006*, pp. 404–417. Springer (2006)
- [14] Behm, D.G., Anderson, K., Curnew, R.S.: Muscle force and activation under stable and unstable conditions. *The Journal of Strength & Conditioning Research* 16(3), 416–422 (2002)
- [15] Bishop, C.M., et al.: *Pattern recognition and machine learning*, vol. 1. springer New York (2006)
- [16] Bouisset, S., Goubel, F.: Interdependence of relations between integrated emg and diverse biomechanical quantities in normal voluntary movements. *Activitas nervosa superior* (1971)

- [17] Breiman, L.: Random forests. *Machine learning* 45(1), 5–32 (2001)
- [18] Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A.: *Classification and regression trees*. CRC press (1984)
- [19] Brubaker, M.A., Sigal, L., Fleet, D.J.: Estimating contact dynamics. In: 2009 IEEE 12th International Conference on Computer Vision. pp. 2389–2396. IEEE (2009)
- [20] Bults, R.G., Knoppel, D.F., Widya, I.A., Schaake, L., Hermens, H.J.: The myofeedback-based tele-treatment system and its evaluation. *Journal of telemedicine and telecare* 16(6), 308–315 (2010)
- [21] Cameron, J.R., Skofronick, J.G., Grant, R.M.: *Physics of the Body*, 2nd Edition. Medical Physics Publishing Corporation (1999)
- [22] Carli, M., Goffredo, M., Schmid, M., Neri, A.: Study of muscular deformation based on surface slope estimation. In: *Electronic Imaging*. pp. 60640U–60640U. International Society for Optics and Photonics (2006)
- [23] Chaffin, D.B., Andersson, G., Martin, B.J., et al.: *Occupational biomechanics*. Wiley New York (1984)
- [24] Chapelle, O., Haffner, P., Vapnik, V.N.: Support vector machines for histogram-based image classification. *Neural Networks, IEEE Transactions on* 10(5), 1055–1064 (1999)
- [25] Chen, N., Urban, S., Bayer, J., Van Der Smagt, P.: Measuring fingertip forces from camera images for random finger poses. In: *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*. pp. 1216–1221. IEEE (2015)
- [26] Chen, N., Urban, S., Osendorfer, C., Bayer, J., Van Der Smagt, P.: Estimating finger grip force from an image of the hand using convolutional neural networks and gaussian processes. In: *Robotics and Automation (ICRA), 2014 IEEE International Conference on*. pp. 3137–3142. IEEE (2014)
- [27] Chitade, A.Z., Katiyar, S.: Colour based image segmentation using k-means clustering. *International J of Engineering Science and Technology* 2(10), 5319–5325 (2010)
- [28] Cholewicki, J., McGill, S.M.: Emg assisted optimization: a hybrid approach for estimating muscle forces in an indeterminate biomechanical model. *Journal of biomechanics* 27(10), 1287–1289 (1994)
- [29] Cholewicki, J., McGill, S.M., Norman, R.W.: Comparison of muscle forces and joint load from an optimization and emg assisted lumbar spine model: towards development of a hybrid approach. *Journal of biomechanics* 28(3), 321–331 (1995)
- [30] Cortes, C., Vapnik, V.: Support-vector networks. *Machine Learning* 20(3), 273–297 (1995)
- [31] Cristina, B.M., Alberto, B., Rita, S., Taian, V.: Integration of surface emg, us imaging and 3d kinematic: New frontiers for muscle function investigation. In: *6th European Conference of the International Federation for Medical and Biological Engineering*. pp. 356–359. Springer (2015)
- [32] Crowninshield, R.D., Brand, R.A.: A physiologically based criterion of muscle force prediction in locomotion. *Journal of biomechanics* 14(11), 793–801 (1981)
- [33] Csurka, G., Dance, C., Fan, L., Willamowski, J., Bray, C.: Visual categorization with bags of keypoints. In: *Workshop on statistical learning in computer vision, ECCV*. vol. 1, pp. 1–2 (2004)
- [34] De Luca, C.J.: The use of surface electromyography in biomechanics. *Journal of applied biomechanics* 13(2), 135–163 (1997)
- [35] De Rugy, A., Loeb, G.E., Carroll, T.J.: Muscle coordination is habitual rather than optimal. *Journal of Neuroscience* 32(21), 7384–7391 (2012)
- [36] De Serres, S.J., Enoka, R.M.: Older adults can maximally activate the biceps brachii muscle by voluntary command. *Journal of Applied Physiology* 84(1), 284–291 (1998)

- [37] Dutta, T.: Evaluation of the kinect sensor for 3-d kinematic measurement in the workplace. *Applied ergonomics* 43(4), 645–649 (2012)
- [38] Edman, K.: The velocity of unloaded shortening and its relation to sarcomere length and isometric force in vertebrate muscle fibres. *J of Physiology* 291(1), 143–159 (1979)
- [39] Enoka, R.M., Duchateau, J.: Muscle fatigue: what, why and how it influences muscle function. *The Journal of physiology* 586(1), 11–23 (2008)
- [40] Erdemir, A., McLean, S., Herzog, W., van den Bogert, A.J.: Model-based estimation of muscle forces exerted during movements. *Clinical biomechanics* 22(2), 131–154 (2007)
- [41] FENGJUN, B.: Muscle force estimation and fatigue detection based on semg signals (2013)
- [42] Goffredo, M., Carli, M., Conforto, S., Bibbo, D., Neri, A., D’Alessio, T.: Evaluation of skin and muscular deformations in a non-rigid motion analysis. In: *Medical Imaging*. pp. 535–541. International Society for Optics and Photonics (2005)
- [43] Goffredo, M., Piras, P., Varano, V., Gabriele, S., D’Anna, C., Conforto, S.: Shape analysis of bicipital contraction by means of rgb-d sensor, parallel transport and trajectory analysis. In: *XIV Mediterranean Conference on Medical and Biological Engineering and Computing 2016*. pp. 634–639. Springer (2016)
- [44] Gordon, A., Huxley, A.F., Julian, F.: The variation in isometric tension with sarcomere length in vertebrate muscle fibres. *J of physiology* 184(1), 170–192 (1966)
- [45] Gorniak, S.L., Zatsiorsky, V.M., Latash, M.L.: Manipulation of a fragile object. *Experimental brain research* 202(2), 413–430 (2010)
- [46] Grieve, T.R., Doyle, C.E., Hollerbach, J.M., Mascaro, S.A.: Calibration of fingernail imaging for multidigit force measurement. In: *Haptics Symposium (HAPTICS), 2014 IEEE*. pp. 623–627. IEEE (2014)
- [47] Grieve, T.R., Hollerbach, J.M., Mascaro, S.A.: Optimizing fingernail imaging calibration for 3d force magnitude prediction. *IEEE transactions on haptics* 9(1), 69–79 (2016)
- [48] Heckathorne, C.W., Childress, D.S.: Relationships of the surface electromyogram to the force, length, velocity, and contraction rate of the cineplastic human biceps1. *American Journal of Physical Medicine & Rehabilitation* 60(1), 1–hyhen (1981)
- [49] Van der Helm, F.C.: A finite element musculoskeletal model of the shoulder mechanism. *Journal of biomechanics* 27(5), 551555–553569 (1994)
- [50] Hislop, H., Avers, D., Brown, M.: *Daniels and Worthingham’s Muscle Testing-E-Book: Techniques of Manual Examination and Performance Testing*. Elsevier Health Sciences (2013)
- [51] Hong, Q.Y., Park, S.I., Hodgins, J.K.: A data-driven segmentation for the shoulder complex. In: *Computer Graphics Forum*. vol. 29, pp. 537–544. Wiley Online Library (2010)
- [52] Hu, M.K.: Visual pattern recognition by moment invariants. *Information Theory, IRE Transactions on* 8(2), 179–187 (1962)
- [53] Johnson, L.A., Fuglevand, A.J.: Evaluation of probabilistic methods to predict muscle activity: implications for neuroprosthetics. *Journal of neural engineering* 6(5), 055008 (2009)
- [54] Katz, S., Tal, A., Basri, R.: Direct visibility of point sets. In: *ACM Transactions on Graphics (TOG)*. vol. 26, p. 24. ACM (2007)
- [55] Kisner, C., Colby, L.A., Borstad, J.: *Therapeutic exercise: foundations and techniques*. Fa Davis (2017)
- [56] Konrad, P.: *The abc of emg. A practical introduction to kinesiological electromyography* 1 (2005)

- [57] Kumar, S.: Biomechanics in ergonomics. CRC Press (1999)
- [58] Laursen, B., Sjøgaard, K., Sjøgaard, G.: Biomechanical model predicting electromyographic activity in three shoulder muscles from 3d kinematics and external forces during cleaning work. *Clinical Biomechanics* 18(4), 287–295 (2003)
- [59] Lawrence, J.H., De Luca, C.: Myoelectric signal versus force relationship in different human muscles. *Journal of Applied Physiology* 54(6), 1653–1659 (1983)
- [60] Lindberg, P., Ody, C., Feydy, A., Maier, M.A.: Precision in isometric precision grip force is reduced in middle-aged adults. *Experimental brain research* 193(2), 213–224 (2009)
- [61] Lippold, O.: The relation between integrated action potentials in a human muscle and its isometric tension. *Journal of physiology* 117(4), 492 (1952)
- [62] Manal, K.T., Buchanan, T.S.: Biomechanics of human movement. *Standard handbook of biomedical engineering & design* p. 26 (2004)
- [63] Matheson Rittenhouse, D., Abdullah, H.A., John Runciman, R., Basir, O.: A neural network model for reconstructing emg signals from eight shoulder muscles: Consequences for rehabilitation robotics and biofeedback. *Journal of biomechanics* 39(10), 1924–1932 (2006)
- [64] McGinnis, P.: Biomechanics of sport and exercise. *Human Kinetics* (2013)
- [65] Meyer, D.: Support vector machines. *The Interface to libsvm in package e1071. e1071 Vignette* (2012)
- [66] Moritani, T., DeVries, H.A.: Reexamination of the relationship between the surface integrated electromyogram (iemg) and force of isometric contraction. *American Journal of Physical Medicine & Rehabilitation* 57(6), 263–277 (1978)
- [67] Mosafavizadeh, M., Wang, L., Lian, Q., Liu, Y., He, J., Li, D., Jin, Z.: Muscle activity prediction using wavelet neural network. In: *Wavelet Analysis and Pattern Recognition (ICWAPR), 2013 International Conference on*. pp. 241–246. IEEE (2013)
- [68] Neumann, T., Varanasi, K., Hasler, N., Wacker, M., Magnor, M., Theobalt, C.: Capture and statistical modeling of arm-muscle deformations. In: *Computer Graphics Forum*. vol. 32, pp. 285–294. Wiley Online Library (2013)
- [69] Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. *Pattern recognition* 29(1), 51–59 (1996)
- [70] Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24(7), 971–987 (2002)
- [71] Palmer, M.L., Epler, M.E., Epler, M.F.: *Fundamentals of musculoskeletal assessment techniques*. Lippincott Williams & Wilkins (1998)
- [72] Pang, G., Neumann, U.: 3d point cloud object detection with multi-view convolutional neural network. In: *Pattern Recognition (ICPR), 2016 23rd International Conference on*. pp. 585–590. IEEE (2016)
- [73] Park, J., Singh, T., Zatsiorsky, V.M., Latash, M.L.: Optimality versus variability: effect of fatigue in multi-finger redundant tasks. *Experimental brain research* 216(4), 591–607 (2012)
- [74] Park, S.I., Hodgins Jessica, K.: Data-driven modeling of skin and muscle deformation. *ACM Transactions on Graphics* 27(3), 96 (2008)
- [75] Perlich, C.: Learning curves in machine learning. In: *Encyclopedia of Machine Learning*, pp. 577–580. Springer (2011)
- [76] Pham, T.H., Kheddar, A., Qammaz, A., Argyros, A.A.: Towards force sensing from vision: Observing hand-object interactions to infer manipulation forces. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 2810–2819 (2015)

- [77] Pham, T., Kyriazis, N., Argyros, A.A., Kheddar, A.: Hand-object contact force estimation from markerless visual tracking (2016), submitted to *IEEE Transactions on Pattern Analysis and Machine Intelligence*
- [78] Potvin, J., Brown, S.: Less is more: high pass filtering, to remove up to 99% of the surface emg signal power, improves emg-based biceps brachii muscle force estimates. *Journal of Electromyography and Kinesiology* 14(3), 389–399 (2004)
- [79] Prentice, S., Patla, A., Stacey, D.: Artificial neural network model for the generation of muscle activation patterns for human locomotion. *Journal of electromyography and kinesiology* 11(1), 19–30 (2001)
- [80] Robertini, N., Neumann, T., Varanasi, K., Theobalt, C.: Capture of arm-muscle deformations using a depth-camera. In: *Proceedings of the 10th European Conference on Visual Media Production*. p. 12. ACM (2013)
- [81] Rogante, M., Silvestri, S., Grigioni, M., Zampolini, M.: Electromyographic audio biofeedback for telerehabilitation in hospital. *Journal of telemedicine and telecare* 16(4), 204–206 (2010)
- [82] Rusu, R.B., Bradski, G., Thibaux, R., Hsu, J.: Fast 3d recognition and pose using the viewpoint feature histogram. In: *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. pp. 2155–2162. IEEE (2010)
- [83] Rusu, R.B., Cousins, S.: 3D is here: Point Cloud Library (PCL). In: *IEEE International Conference on Robotics and Automation (ICRA)*. Shanghai, China (May 9-13 2011)
- [84] Sand, P., McMillan, L., Popović, J.: Continuous capture of skin deformation. *ACM Transactions on Graphics (TOG)* 22(3), 578–586 (2003)
- [85] Sbriccoli, P., Bazzucchi, I., Rosponi, A.e.a., Bernardi, M., De Vito, G., Felici, F.: Amplitude and spectral characteristics of biceps brachii semg depend upon speed of isometric force generation. *Journal of Electromyography and Kinesiology* 13(2), 139–147 (2003)
- [86] Schönauer, C., Pintaric, T., Kaufmann, H., Jansen-Kosterink, S., Vollenbroek-Hutten, M.: Chronic pain rehabilitation with a serious game using multimodal input. In: *Virtual Rehabilitation (ICVR), 2011 International Conference on*. pp. 1–8. IEEE (2011)
- [87] Sekiya, M., Tsuji, T.: Inverse estimation of multiple muscle activations based on linear logistic regression. In: *Rehabilitation Robotics (ICORR), 2017 International Conference on*. pp. 935–940. IEEE (2017)
- [88] Senanayake, S., Triloka, J., Malik, O.A., Iskandar, M.: Artificial neural network based gait patterns identification using neuromuscular signals and soft tissue deformation analysis of lower limbs muscles. In: *International Joint Conference on Neural Networks*. pp. 3503–3510. IEEE (2014)
- [89] Shan, C., Gong, S., McOwan, P.W.: Facial expression recognition based on local binary patterns: A comprehensive study. *Image and Vision Computing* 27(6), 803–816 (2009)
- [90] Sifakis, E., Neverov, I., Fedkiw, R.: Automatic determination of facial muscle activations from sparse motion capture marker data. In: *ACM Transactions on Graphics*. vol. 24, pp. 417–425. ACM (2005)
- [91] Spielholz, P., Silverstein, B., Morgan, M., Checkoway, H., Kaufman, J.: Comparison of self-report, video observation and direct measurement methods for upper extremity musculoskeletal disorder physical risk factors. *Ergonomics* 44(6), 588–613 (2001)
- [92] Staudenmann, D., Roeleveld, K., Stegeman, D.F., Van Dieën, J.H.: Methodological aspects of semg recordings for force estimation—a tutorial and review. *Journal of Electromyography and Kinesiology* 20(3), 375–387 (2010)
- [93] Sural, S., Qian, G., Pramanik, S.: Segmentation and histogram generation using the hsv color space for image retrieval. In: *Image Processing. 2002. Proceedings. 2002 International Conference on*. vol. 2, pp. II–589. IEEE (2002)

- [94] Tibold, R., Fuglevand, A.J.: Prediction of muscle activity during loaded movements of the upper limb. *Journal of neuroengineering and rehabilitation* 12(1), 6–6 (2015)
- [95] Vapnik, V.N., Vapnik, V.: *Statistical learning theory*, vol. 1. Wiley New York (1998)
- [96] Velizhev, A., Shapovalov, R., Schindler, K.: Implicit shape models for object detection in 3d point clouds. In: *International Society of Photogrammetry and Remote Sensing Congress*. vol. 2 (2012)
- [97] Vigouroux, L., Quaine, F., Labarre-Vila, A., Amarantini, D., Moutet, F.: Using emg data to constrain optimization procedure improves finger tendon tension estimations during static fingertip force production. *Journal of biomechanics* 40(13), 2846–2856 (2007)
- [98] Vollenbroek-Hutten, M.M., Hermens, H.J., Kadefors, R., Danneels, L., Nieuwenhuis, L.J., Hasenbring, M.: *Telemedicine services: from idea to implementation* (2010)
- [99] Wohlking, W., Vincze, M.: Ensemble of shape functions for 3d object classification. In: *Robotics and Biomimetics (ROBIO)*, 2011 IEEE International Conference on. pp. 2987–2992. IEEE (2011)
- [100] Woods, J., Bigland-Ritchie, B.: Linear and non-linear surface emg/force relationships in human muscles: An anatomical/functional argument for the existence of both. *American Journal of Physical Medicine & Rehabilitation* 62(6), 287–299 (1983)
- [101] Xu, X., McGorry, R.W., Lin, J.H.: A regression model predicting isometric shoulder muscle activities from arm postures and shoulder joint moments. *Journal of Electromyography and Kinesiology* 24(3), 419–429 (2014)
- [102] Yang, M.H., Ahuja, N.: Gaussian mixture model for human skin color and its applications in image and video databases. In: *Electronic Imaging'99*. pp. 458–466. International Society for Optics and Photonics (1998)
- [103] Yong-Ping, Z., MMF, C., Jun, S., Xin, C., Qing-Hua, H.: Sonomyography: Monitoring morphological changes of forearm muscles in actions with the feasibility for the control of powered prosthesis. *Medical eng. & physics* 28(5), 405–415 (2006)
- [104] Zhao, G., Pietikainen, M.: Dynamic texture recognition using local binary patterns with an application to facial expressions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 29(6), 915–928 (2007)
- [105] Zhao, G., Pietikäinen, M.: Dynamic texture recognition using volume local binary patterns. In: *Dynamical Vision*, pp. 165–177. Springer (2007)
- [106] Zoccolan, D., Giachetti, A., Torre, V.: The use of optical flow to characterize muscle contraction. *Journal of neuroscience methods* 110(1), 65–80 (2001)
- [107] Zoccolan, D., Torre, V.: Using optical flow to characterize sensory-motor interactions in a segment of the medicinal leech. *Journal of Neuroscience* 22(6), 2283–2298 (2002)