SciVerse ScienceDirect
Disponible en www.sciencedirect.com

Revista Iberoamericana de Automática e Informática industrial 00 (2014) 1–10

RIAI
www.elsevier.es/RIAI

Reconocimiento en-línea de acciones humanas basado en patrones de RWE aplicado en ventanas dinámicas de momentos invariantes

Dennis Romero López[a,b,*], Anselmo Frizera Neto[a], Teodiano Freire Bastos[a]

[a] *Departamento de Engenharia Elétrica, Universidade Federal do Espírito Santo, Vitória - Brasil*
[b] *CIDIS - FIEC, Escuela Superior Politécnica del Litoral, Guayaquil - Ecuador*

Resumen

En este trabajo se presenta una metodología para el reconocimiento en-línea de acciones humanas en secuencias de vídeo. Se aborda un enfoque eficiente para el uso de momentos invariantes como descriptores de imagen, aplicados en siluetas obtenidas del procesamiento de mapas de profundidad. Una comparación rápida entre ventanas de tamaño 4 (equivalente a 4 frames) es realizada mediante el cómputo de la distancia de Mahalanobis, sobre una de las secuencias de momentos invariantes identificada como la menos sensible al ruido de captura y la más estable durante ausencia de movimiento. Este enfoque es usado para la detección rápida del estado de parada/movimiento, el cual permite la captura de intervalos (ventanas) de crecimiento dinámico para su posterior procesamiento, rescatando de la señal contenida sus propiedades temporales y frecuenciales. Mediante la aplicación de la transformada Wavelet Haar, tres niveles de descomposición son utilizados para el cómputo de la Energía Relativa Wavelet (RWE - Relative Wavelet Energy) y SSC (Slope Sign Change), obteniendo patrones 11-dimensionales. En experimentos realizados, el 97 % de 4 movimientos capturados en-línea fueron reconocidos correctamente, y 10 movimientos tomados de la base de datos Muhavi-MAS fueron reconocidos con 94,2 % de efectividad.

Palabras Clave:
Visión por ordenador, Mapas de profundidad, Reconocimiento de acciones humanas, Relative Wavelet Energy, Distancia de Mahalanobis.

1. Introducción

La identificación y análisis automático del movimiento humano ha atraído interés durante varias décadas (Poppe, 2010; Moeslund and Kruger, 2006; Wang et al., 2003), y continúa siendo un área activa de investigación en el campo de visión por ordenador y procesamiento de señales. Nuevos enfoques están dando lugar a adelantos en temas relacionados con reconocimiento de expresiones humanas, orientado a robótica e interacción hombre-máquina, asistencia de movimiento en prótesis biomecánicas, descripción semántica de movimiento y retroalimentación automática durante Fisioterapia. Con respecto a la rehabilitación física, el análisis cualitativo y cuantitativo de patrones de movimiento es esencial para monitorizar la recuperación funcional de pacientes durante rehabilitación, y los métodos empleados deben ser lo suficientemente flexibles para permitir una amplia diversidad de aplicaciones clínicas (Salas-Lopez et al., 2012). Adicionalmente, la creciente capacidad de procesamiento de dispositivos portables posibilitan el planteamiento de soluciones a problemas cada vez más complejos, extendiendo el número de aplicaciones que requieren de procesamiento visual o sirviendo como proveedores de información de entrada para otros dispositivos y aplicaciones, como lo propuesto en (Garcia-Costa et al., 2011), donde se asumen sistemas embarcados de comunicación para evitar colisiones de vehículos en fila.

1.1. Especificación del problema

En este trabajo se presenta un enfoque para el problema de reconocimiento en-línea de movimientos de personas. La identificación de acciones humanas usando técnicas de visión por ordenador involucra el abordaje de diferentes sub-problemas, como la detección/localización de la persona, extracción y seguimiento de características, creación de patrones de movimiento y clasificación. La segmentación robusta de imágenes es un problema complejo, debido a la variabilidad de los parámetros que conforman la escena de vídeo (posición del objeto respecto

*Autor en correspondencia
Correos electrónicos: dennis@ele.ufes.br (Dennis Romero López), anselmo@ele.ufe.br (Anselmo Frizera Neto), tfbastos@ele.ufes.br (Teodiano Freire Bastos)

a la cámara, cambios de iluminación, resolución de la imagen, etc.). Una solución única presentaría dificultades en la adaptación de los parámetros de segmentación si los objetivos de la aplicación varían. La metodología propuesta hace uso de mapas de profundidad, obtenidos de un sensor RGB-D (Kinect), el cual incluye emisores de infrarrojo para aportar información de profundidad, contribuyendo de esta forma al proceso de segmentación. Mapas de profundidad han sido utilizados en trabajos recientes relacionados con detección de peatones, como los propuestos en (Broggi et al., 2000) y (Soga et al., 2005). En este trabajo, el procesamiento de mapas de profundidad capturados en ambientes internos, dio como resultado siluetas de gran calidad, las cuales pueden ser obtenidas incluso en total carencia de iluminación.

La extracción de características para seguimiento es otro de los problemas a considerar, el cual forma parte de los principales retos en el campo de visión por ordenador. Esto es debido a que el movimiento de personas en secuencias de vídeo involucra principalmente variaciones de escala y traslación. Aunque existen métodos que han sido ampliamente utilizados por ser eficientes y bastante robustos ante variaciones de escala, rotación y traslación, como los citados en (Chen et al., 2004), (Huang and Leng, 2010) y (Hu et al., 2007), en algunos casos la cantidad de características a extraer da lugar al manejo de altas dimensiones, haciendo necesario el uso posterior de algoritmos de reducción de dimensionalidad (Kernel PCA, Sammon maps, LLP, LLE, etc.). Entre los métodos mencionados de reducción, se encuentran aquellos que son extensivamente iterativos (Yan et al., 2011), dificultando su utilización en aplicaciones con requerimientos rápidos de respuesta. En este trabajo se propone un enfoque diferente para el uso de momentos invariantes de Hu (Huang and Leng, 2010), método que ha sido ampliamente citado en la literatura para el reconocimiento de imágenes, y de los mejores en términos de su ortogonalidad, invarianza a rotación y computación rápida (Gonzalez, 2010; Chen et al., 2004). Sin embargo, su sensibilidad al ruido en la imagen los hacen poco robustos al ser aplicados directamente sobre imágenes sin un extensivo pre-procesamiento.

Dado que este trabajo toma mapas de profundidad como datos de entrada, los momentos invariantes de Hu son aplicados sobre las siluetas obtenidas, reduciendo la presencia de ruido en la señal resultante. Adicionalmente, son considerados exclusivamente tres de los momentos computables, obteniendo descriptores de silueta con un costo computacional menor al tiempo entre frames (para un ordenador multi-core de 2GHz), lo cual posibilita el aprovechamiento de todos los frames provistos por el sensor de captura, mejorando la resolución para los métodos posteriores de extracción de patrones de movimiento y detección del estado de parada/movimiento.

La búsqueda de patrones que permitan la identificación automática de movimientos es otro de los problemas ampliamente abordados. Métodos para la extracción de información relevante mediante procesamiento digital de señales han sido estudiados en diferentes áreas, los cuales abordan el movimiento humano con señales de EMG (Electromiografía), sensores inerciales (IMU - Inertial Measurement Unit), métodos basados en visión por ordenador, entre otros (Cifuentes et al., 2012).

Uno de los enfoques más comunes para el estudio de señales, es el análisis frecuencial por medio de la Transformada Rápida de Fourier (FFT), el cual ha sido abordado en etapas previas al presente trabajo (Romero et al., 2012a), lo que ha permitido entender más profundamente el comportamiento de los momentos invariantes de Hu durante el seguimiento de actividades humanas. Sin embargo, el movimiento de personas incluye información temporal valiosa, la cual es desconsiderada con abordajes estrictamente frecuenciales. Con el objetivo de extraer tanto información frecuencial como temporal de las señales de movimiento, este trabajo presenta un enfoque basado en transformada Wavelet Haar.

Como será detallado en secciones posteriores, nueve características extraídas de la transformada Wavelet y dos características locales dan como resultado patrones adecuados para clasificación, usando redes neuronales artificiales o métodos probabilísticos, entre los que fueron comparados: RBF (Radial Basic Function), SVM (Support Vector Machines) y K-NN (k-Nearest Neighbor). Varias familias Wavelet fueron estudiadas y comparadas durante el proceso de extracción de características de movimiento, entre ellas Wavelets Daubechies, obteniendo mejores resultados con la transformada Wavelet Haar (sección 2.4).

1.2. Trabajos relacionados

Diferentes aportes en la literatura han permitido nuevos avances en áreas de visión por ordenador, procesamiento digital de señales y reconocimiento de patrones. Esto ha dado lugar a soluciones cada vez más eficientes y robustas para aplicaciones específicas, como por ejemplo: Sistemas de Protección de Peatones, Fisioterapia Asistida por Ordenador, Análisis Semántico de Movimientos, Robótica Asistencial, Evaluación Remota de Pacientes con Déficit Motor, Interfaces Naturales Hombre-Máquina, entre otros. En la revisión del estado del arte fueron identificados trabajos recientes en el área, como los mencionados en (Chockalingam et al., 2009; Park and Trivedi, 2008; Geronimo et al., 2010). Inicialmente, se estudiaron métodos para analizar movimientos de personas en secuencias de vídeo y fue tratado el problema de la extracción de características en diferentes actividades en tiempo, tales como: caminar, saltar, correr, etc. Sin embargo, la diversidad y complejidad de los movimientos humanos llevó a buscar soluciones menos restrictivas a movimientos específicos y a extraer patrones característicos que representen de mejor manera los movimientos encontrados. Considerando los diversos retos dentro del área de análisis visual automático de movimientos, han sido destacados de la literatura trabajos relacionados en cada una de las etapas de pre-procesamiento, segmentación, detección de la persona, seguimiento y clasificación.

Aunque el procesamiento a bajo nivel generalmente es poco mencionado, como en el caso de los ajustes de exposición y rango dinámico, algunos trabajos recientemente publicados consideran conveniente el realzado inicial de imágenes (Marsi et al., 2007), (Knoll, 2007), (Nayar and Branzoi, 2003). El realce de imágenes en tiempo real es una tarea difícil, especialmente en escenarios urbanos, tales como túneles cortos, calles

angostas y movimientos rápidos en la escena, condiciones comunes en aplicaciones de detección de peatones, por ejemplo. En este ámbito, (Nayar and Branzoi, 2003) presentan un enfoque para rango dinámico adaptativo mediante la fusión de diferentes valores de exposición, filtros espaciales, múltiples sensores imagen/pixel y exposición por pixel, etc.

Como menciona la literatura, el análisis basado en estéreo visión para segmentación de objetos en la imagen es la opción que ha tenido mayor éxito. Trabajos recientes en este tema han demostrado buenos resultados, tal como los presentados por (Franke and Joos, 2000; Grubb et al., 2004). Por otro lado, el análisis basado en dos dimensiones ha tenido pocos resultados trascendentes en esta etapa. La variabilidad de los parámetros que conforman la escena de vídeo hace frecuentemente necesario el uso de información adicional, como por ejemplo, mediciones de profundidad que pueden ser provistos por otros dispositivos. Posteriormente, en las etapas de detección y seguimiento de la persona, métodos recientes presentan diferentes abordajes, como en (Chockalingam et al., 2009), donde se propone un método que divide el objeto en varios fragmentos o regiones, los cuales son representados por un Modelo de Mistura de Gaussianas (GMM) en un espacio de características espaciales agrupadas. El modelado del objeto y del fondo son realizados conforme el algoritmo de Chan-Vese (Chan and Vese, 2001), y las regiones resultantes son usadas para aprender la forma dinámica del objeto en el tiempo, manteniendo el seguimiento continuo hasta en casos de oclusión total.

Trivedi y otros presentan en (Park and Trivedi, 2007) un enfoque sinérgico del cuerpo de la persona, basado en jerarquía de acciones, considerando pose estática, gestos dinámicos y acción de partes del cuerpo durante actividades de la persona, usando los algoritmos de Baum-Welch y Viterbi (Rabiner, 1989) para codificar HMMs independientes. El mismo autor introduce en (Park and Trivedi, 2008) el concepto de espacio personal tiempo-espacio para direccionar diferentes comportamientos de personas. En este tema, nuestro trabajo propone un método para la extracción de patrones en secuencias de características, las cuales pueden ser obtenidas de diferentes métodos de seguimiento de personas, como los mencionados anteriormente.

Por otro lado, estudios basados en momentos invariantes y transformada Wavelet fueron realizados por (Sarvaiya, 2011), para registro automático de imágenes, aplicados en correspondencia de dos o más imágenes diferentes. El método estima los parámetros del modelo de transformación geométrica, que realiza un mapeamiento inverso de las imágenes capturadas hasta su imagen referencia. Puntos característicos de ambas imágenes son extraídos usando Wavelet Sombrero Mejicano y la correspondencia de puntos es seguida con momentos invariantes. Las propiedades de la transformada Wavelet han sido reconocidas y sus aplicaciones son muy variadas. Estudios recientes en el área utilizan características Wavelet como en (Viola et al., 2003; Jones and Snow, 2008) por medio de Wavelets Haar. En otros campos de investigación, como es el área de Bioingeniería, el trabajo presentado en (Haibin et al., 2008) calcula la Energía Relativa Wavelet para extraer características útiles en el diseño de una interfaz cerebro-ordenador, mediante el análisis de señales de EEG (Eletroencefalografía) para controlar programas de ordenador y otros dispositivos como silla de ruedas, etc. Los enfoques bioinspirados están siendo considerados en soluciones basadas en visión por ordenador, orientados al estudio de movimientos humanos como las propuestas de Miau y otros en (Miau et al., 2001) y (Itti et al., 1998).

El presente trabajo da continuidad al estudio presentado en (Romero et al., 2012b), mostrando en esta versión mejoras significativas respecto a la propuesta anterior, las cuales son detalladas a continuación:

- Extracción de características Wavelet Haar en lugar de Daubechies 4, basado en nueva evidencia de resultados y tendencias encontradas en la revisión de estado del arte.
- Adaptación de los parámetros posteriores a la extracción de momentos invariantes (normalización de potencias y ajuste del umbral de distancia de Mahalanobis), logrando de esta forma reducir el tamaño de la ventana de análisis temporal, mejorando la sensibilidad y el tiempo de detección de movimientos.
- Incorporación del método de análisis de derivada (SSC), incrementando en dos el número de características del patrón resultante, obteniendo de esta forma una mayor separación de los patrones, disminución del número de repeticiones para entrenamiento y el soporte de una mayor cantidad de movimientos. Resultados comparativos son mostrados en la sección 3.

1.3. Generalidades del enfoque propuesto

En este trabajo se presenta una metodología para el reconocimiento en-línea de acciones humanas en secuencias de vídeo. El método incluye un enfoque eficiente para el uso de momentos invariantes como descriptores de imagen (Gonzalez, 2010), aplicados en siluetas obtenidas del procesamiento de mapas de profundidad. Una comparación rápida entre ventanas de 4 muestras (equivalente a 4 frames) es realizada mediante el cómputo de la distancia de Mahalanobis, sobre una de las secuencias de momentos invariantes, identificada como la menos sensible al ruido de captura y la más estable durante ausencia de movimiento. Este enfoque es usado para la detección rápida del estado de parada/movimiento, el cual permite la captura de intervalos (ventanas) de crecimiento dinámico para su posterior procesamiento, rescatando de la señal contenida sus propiedades temporales y frecuenciales. Mediante la aplicación de la transformada Wavelet Haar, 3 niveles de descomposición son utilizados para el cómputo de la Energía Relativa Wavelet (RWE - Relative Wavelet Energy)(Rosso et al., 2003) y SSC (Slope Sign Change) (Phinyomark et al., 2009), obteniendo patrones 11-dimensionales para su clasificación usando redes neuronales. La Figura 1 muestra el diagrama de la metodología propuesta.

2. Metodología

La presente propuesta forma parte de estudios orientados al entendimiento de la dinámica humana, dentro del campo de

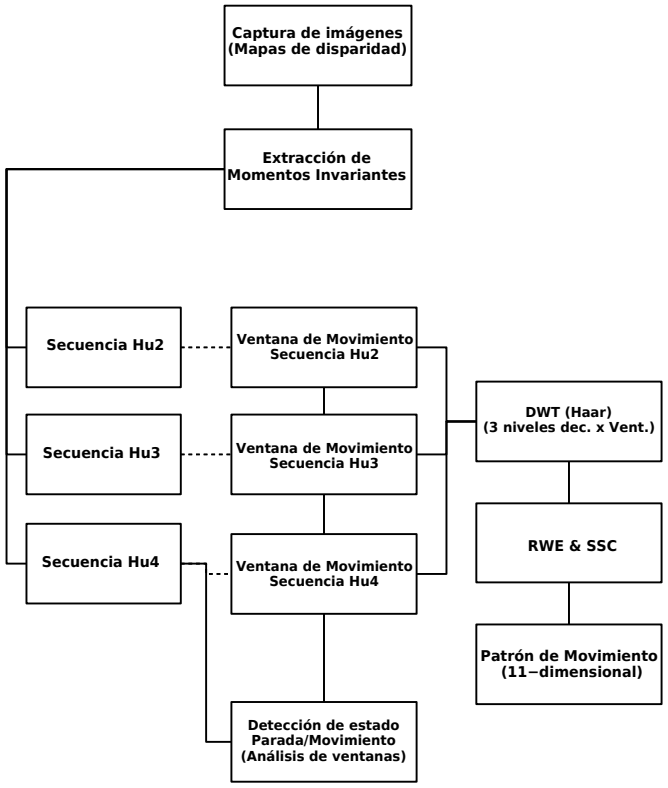

Figura 1: Diagrama de la metodología propuesta.

interacción hombre-máquina o más específicamente hombre-robot. Teniendo como consideración principal la interacción desde un punto de vista natural, esta propuesta aborda la identificación automática de movimientos como el proceso posterior a la localización/segmentación de la persona, por lo que dentro del campo objetivo de esta propuesta, se busca aportar dentro de los siguientes propósitos específicos:

- Extracción en tiempo real de características en cada imagen, que de lugar con posterior procesamiento, a la identificación de expresiones corporales del cotidiano, gestos, señales universales y otros movimientos no comunes, mediante aprendizaje supervisado.
- Identificación del inicio y fin de un movimiento relevante durante una expresión corporal compleja, mediante la detección rápida (usando una mínima cantidad de imágenes) de parada y movimiento.
- Extracción de patrones que permitan no sólo el reconocimiento de gestos o expresiones aprendidas, sino también la medición y comparación con otros movimientos de interés, útiles en diferentes áreas de aplicación, como por ejemplo, robótica de rehabilitación, entre otras.

En función de los objetivos mencionados, la metodología a seguir describe inicialmente el conjunto de datos de entrada, representados por mapas de profundidad capturados del sensor Kinect, para la posterior generación de siluetas (Figura 2). De estas siluetas se obtienen descriptores globales como se detalla en la sección 2.1. La sección 2.2 explica el enfoque usado para la detección en-línea del estado de parada/movimiento y la subsecuente extracción de ventanas dinámicas en la sección 2.3. Adicionalmente, en la sección 2.4 se detalla el criterio usado para la extracción de patrones de movimiento. Finalmente, se presentan los resultados experimentales y conclusiones en las secciones 3 y 4, respectivamente.

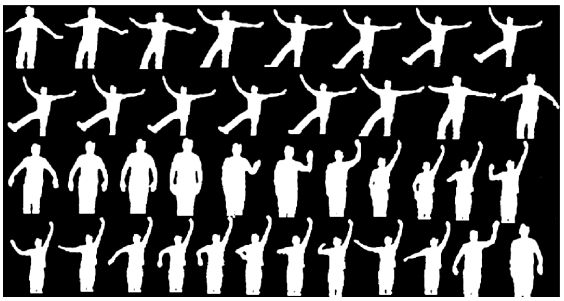

Figura 2: Siluetas obtenidas del procesamiento de mapas de profundidad.

2.1. Características invariantes de silueta por imagen

Para la identificación de patrones de movimiento de cada una de las imágenes, se obtuvo el segundo, tercero y cuarto momento de Hu. El primer momento de Hu fue descartado por presentar redundancias con otros momentos, y por ser más sensible a ruidos en la imagen, mientras que los últimos (quinto, sexto y séptimo momentos) no fueron considerados por presentar valores de orden muy inferior en ciertas condiciones de movimiento. La Figura 3 muestra los siete momentos de Hu normalizados, extraídos durante la realización de 32 movimientos diferentes, entre los que se encuentran gestos de saludo, señales de alerta, señas indicativas y otras expresiones corporales como las ilustradas en la Figura 2. Es posible observar en la Figura 3 los momentos cinco y siete, los cuales presentan valores inferiores a 10^{-14}. El sexto momento de Hu aporta poca información en determinados movimientos como se observa en la figura.

Con el objetivo de utilizar los siete momentos de Hu como características de imagen, algunos autores normalizan las potencias de los valores obtenidos de las siluetas, como lo propuesto en (Mercimek et al., 2005). Sin embargo, en el presente trabajo además de la información relevante de movimiento, se toma en cuenta también el procesamiento en-línea de las señales resultantes (detección de parada/movimiento, extracción de características a partir de las ventanas dinámicas y reconocimiento de patrones). Por tal motivo, se propone el uso de solo tres momentos (Hu2, Hu3 y Hu4) con potencias unificadas, y realizar una normalización [0-1] únicamente en las ventanas dinámicas, anterior a la extracción de patrones Wavelet, como es mostrado en la Figura 5.

En estudios previos (Romero et al., 2012a), fue realizado un análisis estocástico de los momentos invariantes de Hu, donde fue posible identificar características estacionarias después de aplicada la primera derivada, detallada en (Antoniou, 2005). Sin embargo, con el objetivo de conservar información y facilitar la identificación de variaciones en la secuencia de momentos,

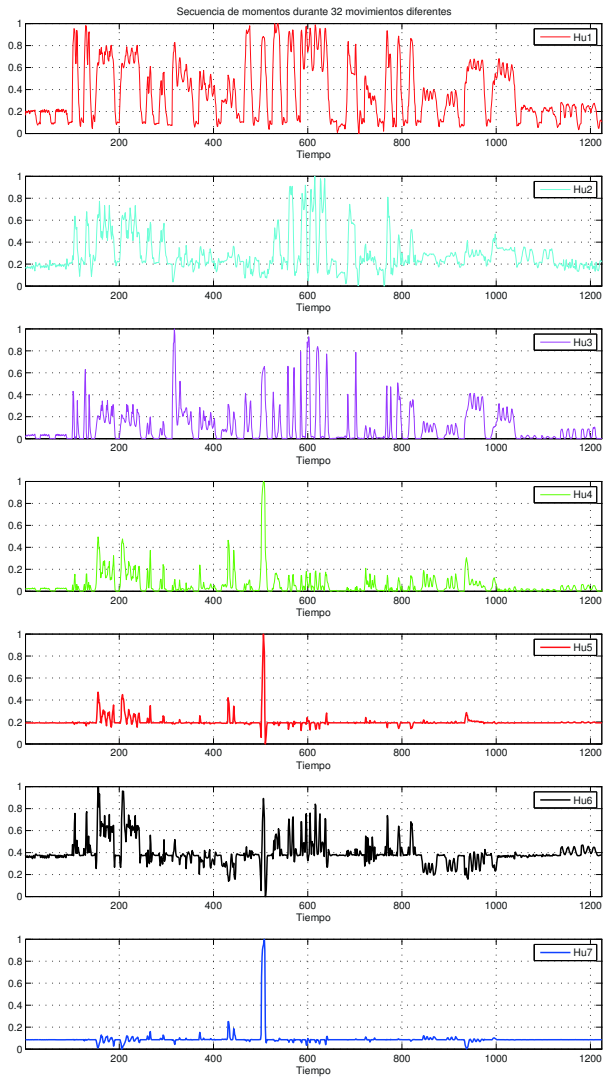

Figura 3: Siete momentos de Hu normalizados, extraídos de las siluetas obtenidas.

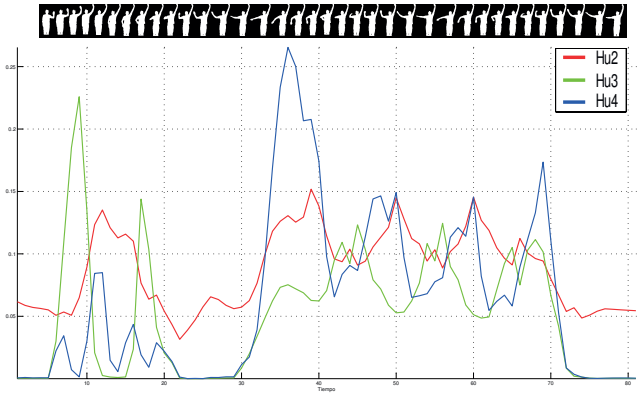

Figura 4: Momentos Hu2, Hu3 y Hu4 durante el movimiento de levantar brazo izquierdo y realizar movimiento circular con el derecho.

así como reducir el costo computacional, la aplicación de la primera derivada y posteriores abordajes únicamente frecuenciales fueron desconsiderados. Las variaciones mencionadas incluyen el incremento o decremento proporcional de los momentos de Hu con movimientos de la parte superior o inferior del cuerpo, respectivamente.

2.2. *Identificación del estado de parada/movimiento*

Los momentos invariantes extraídos para cada imagen permiten representar posturas de una persona en instantes de tiempo de su movimiento. Cada postura se encuentra representada por $Pos_{\delta t}$ y descrita en un espacio de tres dimensiones $Pos_{\delta t} = [\phi_2, \phi_3, \phi_4]$, donde ϕ_i corresponde a los momentos de Hu 2, 3 y 4. En este trabajo, para un intervalo de tiempo τ, se considera cada dimensión como una señal (S) unidimensional, donde $S_i = [\phi_{i0}, \phi_{i1}, \phi_{i2}, \phi_{i3}, \ldots, \phi_{i\tau-1}]$, para $i = \{2, 3, 4\}$. El criterio de análisis unidimensional es usado para identificar el estado de parada/movimiento, donde S_4 es considerado para detectar instantes de inicio y final del movimiento de la persona, por presentar mayor sensibilidad en presencia de movimientos. La Figura 4 muestra los tres momentos seleccionados durante la realización de un gesto de levantar el brazo izquierdo y realizar movimiento circular con el derecho.

Dado que las secuencias de momentos invariantes es actualizada a una tasa aproximada a la capacidad de la cámara de capturar imágenes (30 fps), fue considerado el análisis de ventanas temporales para identificar el inicio y fin de una acción o movimiento de la persona. El tamaño de la ventana fue definido mediante la optimización de la distancia de Mahalanobis (1), la cual puede ser definida como una medida de semejanza entre dos vectores aleatorios $\vec{W_{i-1}}$ y $\vec{W_i}$ con la misma distribución, y con matriz de covarianza Σ.

$$d(\vec{W_{i-1}}, \vec{W_i}) = \sqrt{(\vec{W_{i-1}} - \vec{W_i})^T \Sigma^{-1} (\vec{W_{i-1}} - \vec{W_i})} \quad (1)$$

La distancia de Mahalanobis ha sido utilizada en diversas aplicaciones, tales como en (Qiao et al., 2011), donde fue usada para el reconocimiento de firma. En la Figura 3 es posible observar un cuarto momento de Hu bastante estable en condiciones de carencia de movimiento, por lo que la distancia en estos instantes de tiempo pueden tomar valores inferiores a 0.007, en condiciones de iluminación controlada. Sin embargo, el parámetro de distancia tolerable para la identificación de estados de parada/movimiento podría variar dependiendo de las condiciones de iluminación o presencia de ruido en el proceso de captura de la imagen.

Dada la naturaleza del sensor de captura (Kinect), la calidad de las siluetas obtenidas de los mapas de profundidad capturados varía en función de sus limitaciones, especialmente la sensibilidad a condiciones de alta luminosidad (ambientes externos). Bajo estas condiciones el sensor pierde precisión, incrementando el ruido en las señales de momentos invariantes, comprometiendo de esta forma la efectividad de la segmentación de movimientos y posterior generación de patrones. Debido a esto, los experimentos realizados en este trabajo han sido realizados en ambiente interno. Otros inconvenientes como el limitado campo de visión, han sido superficialmente abordados

mediante la construcción de una base giratoria sobre la que el sensor va montado, con el objetivo de acompañar el desplazamiento de la persona. Sin embargo, la metodología propuesta en este trabajo no busca estar limitado al uso de Kinect, por lo que la solución giratoria mencionada y pruebas en ambientes externos no fueron considerados en este artículo.

En pruebas realizadas en ambiente interno se estableció el valor de distancia como un umbral constante de 0.009. Ventanas de tamaño 4 fueron suficientes para establecer mediciones de distancia que soporten los ruidos del proceso de captura y binarización, proporcionando una marca significativa de inicio y final del movimiento.

Con base en lo expuesto, si el valor de distancia de Mahalanobis entre la ventana temporal (vector de tamaño 4, conteniendo valores del cuarto momento de Hu) en tiempo $\tau - 1$ y la ventana temporal actual en τ es mayor que el umbral establecido, se da inicio a la captura de información correspondiente a las ventanas de movimiento. Estas ventanas de movimiento crecerán dinámicamente hasta que el valor de distancia de las ventanas temporales vuelva a valores por debajo del umbral, marcando de esta forma el final del movimiento desarrollado por la persona.

El uso de características de imagen invariantes a escala y traslación facilita la captura de información a diferentes distancias entre la persona y el sensor. Por otro lado, la velocidad con la que el individuo realiza un movimiento deja de ser relevante dentro de las ventanas dinámicas, es decir, entre las marcas de inicio y final de movimiento. Esto se debe principalmente al posterior procesamiento, con el cual se soportan variaciones de velocidad considerando los siguientes aspectos: 1) La transformada Wavelet Haar, la cual conserva tanto información temporal como frecuencial de las señales de movimiento. 2) El análisis de SSC, el cual refuerza información frecuencial y da un indicio de la forma de la señal. Adicionalmente, se asume que un movimiento determinado será desarrollado de forma diferente entre individuos, incluso entre un mismo individuo. Por tal razón, el uso de redes neuronales permite al sistema "aprender" a tolerar diferencias entre movimientos de la misma naturaleza.

El método de distancia de Mahalanobis fue comparado en estudios previos con el criterio de autocorrelación entre ventanas, obteniendo mediciones más estables con Mahalanobis. Las ventanas de tamaño 4 permiten identificar un estado de parada/movimiento en $tam_{vent}/fps \approx 0{,}2s$, siendo tam_{vent} el tamaño de la ventana temporal.

2.3. *Ventanas dinámicas de movimiento*

Una vez identificado el inicio del movimiento (el cual fue determinado con secuencias del cuarto momento de Hu), los nuevos vectores de momentos son almacenados temporalmente en ventanas de crecimiento dinámico, entre la marca de inicio y final del movimiento $t_{act} - tam_{vent}$. Esto permite incluir en la ventana de movimiento los frames que fueron usados en la detección de inicio del movimiento.

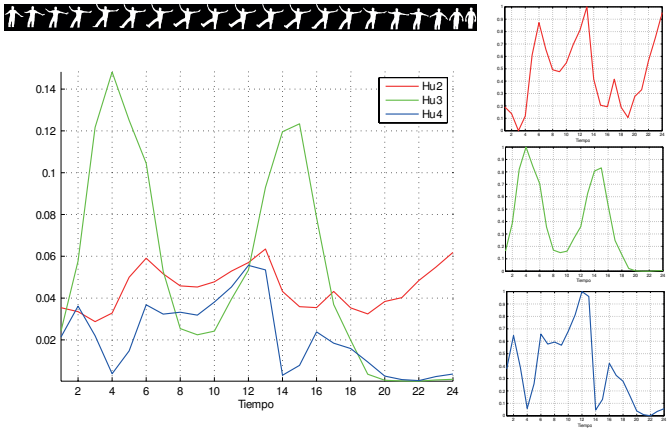

Figura 5: Ventanas de movimiento (izq.) y valores normalizados (der.), extraídos durante la realización de una expresión corporal.

2.4. *Patrones de movimiento*

Debido a la naturaleza periódica de ciertos movimientos humanos (caminar, correr, saltar, etc.), en estudios iniciales fue considerado el análisis frecuencial mediante Transformada Rápida de Fourier (FFT), después de aplicada la primera derivada a las secuencias de momentos invariantes. Sin embargo, la transformada Wavelet discreta (DWT) específicamente permite la discriminación de señales no-estacionarias con diferentes características de frecuencia. En estudios previos (Romero et al., 2012b) fue considerada la DWT con familia Daubechies (db4), obteniendo mejores resultados con Wavelet Haar, por lo que fue escogido como método para la posterior extracción de características de los movimientos de la persona.

La transformada Wavelet discreta es una transformación de la señal temporal original en un espacio de base Wavelet. La representación tiempo-frecuencia Wavelet es realizada filtrando repetidamente una dupla de filtros que dividen el dominio de la frecuencia. Específicamente, la DWT descompone la señal original en una señal de aproximación y una de detalle. La señal de aproximación es subsecuentemente dividida en nuevas señales de aproximación y detalle. Este proceso es repetido iterativamente, produciendo un conjunto de señales de aproximación en diferentes niveles de detalle, con una aproximación final de la señal.

El detalle D_j y la aproximación A_j en el nivel j pueden ser obtenidos filtrando la señal con un filtro pasa alto L-muestras g, y un filtro pasa bajo L-muestras h. Ambas señales de aproximación y detalle son sub-muestreadas por un factor de 2. Esto puede ser expresado conforme a las ecuaciones (2) y (3).

$$A_j[n] = \mathbf{H}\langle A_{j-1}[n]\rangle = \sum_{k=0}^{L-1} h[k]A_{j-1}[2n-k], \tag{2}$$

$$D_j[n] = \mathbf{G}\langle D_{j-1}[n]\rangle = \sum_{k=0}^{L-1} g[k]A_{j-1}[2n-k], \tag{3}$$

donde $A_0[n], n = 0, 1, \ldots, N-1$ es la secuencia temporal original, $\mathbf{H}$ y $\mathbf{G}$ representan los operadores de convolución/submuestreo.

Las secuencias $g[n]$ y $h[n]$ son asociadas con la función Wavelet $\psi(t)$ y la función de escala $\varphi(t)$ a través del producto interno (4) y (5).

$$g[n] = \langle \psi(t),\ \sqrt{2}\psi(2t-n) \rangle, \tag{4}$$

$$h[n] = \langle \varphi(t),\ \sqrt{2}\varphi(2t-n) \rangle. \tag{5}$$

Como fue mencionado anteriormente, la transformada Wavelet $\psi(t)$ seleccionada es Haar, con 3 niveles de descomposición, y puesto que esta representa una base ortogonal para L el concepto de energía está relacionado con las notaciones usuales derivadas de la teoría de Fourier. Entonces, los coeficientes Wavelet están definidos por $C_j(k) = \langle Vent_{mov}, \psi_{j,k} \rangle$, los cuales pueden ser interpretados como los errores locales residuales entre señales de aproximación sucesivas en las escalas j y $j+1$, siendo la energía en cada nivel de descomposición $j = -1, \ldots, -N$ la energía de la señal de detalle (Rosso et al., 2001), definido en (6).

$$E_j = \|r_j\|^2 = \sum_k |C_j(k)|^2, \tag{6}$$

donde $r_j(t)$ es la señal residual en la escala j, y la energía en el instante k es dada por (7).

$$E(k) = \sum_{j=-N}^{-1} |C_j(k)|^2 \tag{7}$$

Como consecuencia, la energía total puede ser obtenida por (8).

$$E_{total} = \|Vent_{mov}\|^2 = \sum_{j<0} \sum_k |C_j(k)|^2 = \sum_{j<0} E_j \tag{8}$$

Finalmente, son definidos los valores p_j normalizados (9), los cuales representan la energía relativa Wavelet, donde $\sum_j p_j = 1$ y la distribución p_j puede ser considerada como una densidad tiempo-escala, por lo cual el método constituye una herramienta adecuada para la detección y caracterización de fenómenos específicos en planos de tiempo y frecuencia.

$$p_j = \frac{E_j}{E_{total}} \tag{9}$$

La energía relativa de los tres niveles de descomposición es calculada para cada ventana de movimiento, lo que resulta en un patrón 9-dimensional, sobre los que se adicionan 2 patrones basados en Slope Sign Change (10).

Zero crossing (ZC) es el número de veces que el valor de amplitud de la señal cruza el eje-Y en su valor de 0. Este método fue utilizado en (Phinyomark et al., 2009), usando también un umbral constante para evitar ruidos en señales de EMG. Para este trabajo, se consideró el enfoque de Slope Sign Change (SSC) que es similar a ZC. Este es otro método que representa información de frecuencia de la señal y representa el número de cambios entre picos positivos y negativos a lo largo de cada una de las tres señales de movimiento capturadas (10).

$$SSC = \sum_{n=2}^{N-1} f[(x_n - x_{n-1})(x_n - x_{n+1})]; \tag{10}$$

$$\mathrm{f(x)} = \begin{cases} 1, & if\, x > umbral \\ 0, & otros \end{cases}$$

3. Resultados y Discusión

El uso de mapas de profundidad ayudó a resolver varios de los problemas mencionados en la revisión bibliográfica acerca del pre-procesamiento y segmentación de la imagen. La información de profundidad provista por el sensor utilizado permitió la creación de siluetas limpias (sin ruidos de fondo) tolerante a la presencia de objetos dinámicos en la escena. Aunque el uso de emisores de infrarrojo (IR) presenta desventajas en ambientes externos con alta luminosidad, los resultados obtenidos en ambientes internos demuestran el potencial de dispositivos que fusionan información de cámaras de vídeo con otros sensores. Un estudio comparativo de diferentes enfoques basados en la fusión de sensores fue realizado por Sappa y otros en (Geronimo et al., 2010) orientado a sistemas de protección de peatones.

Es válido mencionar que el surgimiento de nuevos dispositivos que combinan visión estéreo e información de profundidad ayudarán a reducir las limitaciones de iluminación y distancia de los sensores actuales. Asimismo, estos han demostrado ser útiles en aplicaciones de visión por ordenador. En esta propuesta, momentos de Hu han sido empleados, tomando ventaja de la favorable calidad de las siluetas obtenidas mediante el uso de información adicional de profundidad, lo que permitió aprovechar sus propiedades de invarianza y eficiencia computacional. Sin embargo, existen diversos métodos para la extracción de características invariantes, por lo que el enfoque propuesto podría ser aplicado a secuencias de otro tipo de características en el tiempo.

El análisis temporal basado en distancia de Mahalanobis permitió la detección rápida del estado de parada/movimiento (aprox. 0.2 s), posibilitando la obtención de ventanas de crecimiento dinámico conteniendo sólo información de movimiento. Estas ventanas de movimiento son directamente procesadas para la extracción posterior de características, mediante el computo de Energía Relativa Wavelet. El método de extracción de características basado en transformada Wavelet Haar tuvo un significativo rendimiento computacional, considerando que el tiempo de procesamiento para movimientos cortos (gestos, o expresiones corporales de corta duración) es de aproximadamente 2 ms (ordenador multi-core de 2GHz). Las Figuras 6 y 7 muestran una captura de pantalla de la aplicación desarrollada con los métodos detallados, para la identificación en-línea de movimientos de diferente duración.

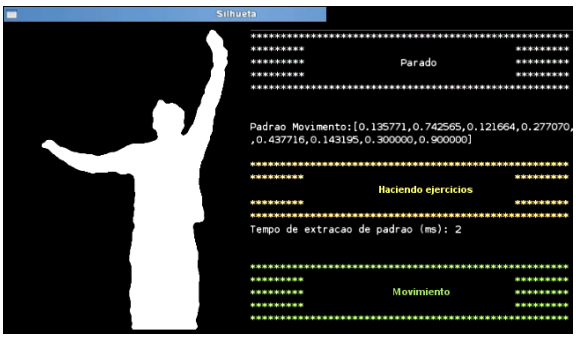

Figura 6: Captura de pantalla durante el reconocimiento de movimientos.

Figura 7: Captura de pantalla durante la realización de un gesto indicativo.

Como fue mencionado inicialmente, las características extraídas de la silueta de una persona conforman señales unidimensionales, donde una de ellas (Hu4) es analizada por el método de identificación de parada/movimiento. Estas señales pueden ser sustituidas por otras señales que describan el movimiento de la persona. De esta forma, el método de extracción de características no depende exclusivamente de los momentos invariantes de la imagen, siendo posible la incorporación de información de movimiento obtenido de sensores inerciales, por ejemplo. Otra de las ventajas del método de extracción de patrones es que este no depende de la creación de modelos de transición, como los citados en (Park and Trivedi, 2008), los cuales limitan el número de movimientos a reconocer.

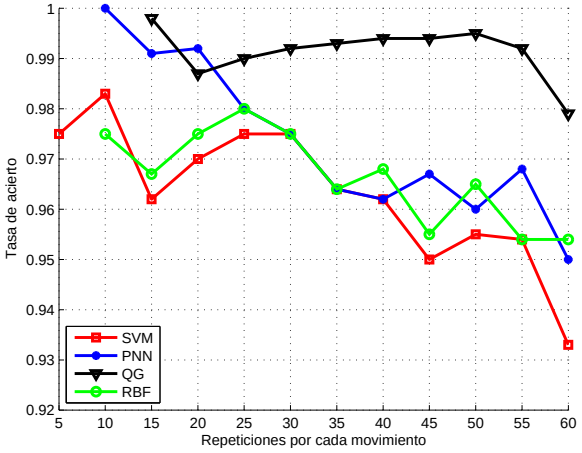

Figura 8: Resultado de la clasificación de los patrones en función de la cantidad de repeticiones por movimiento.

La Figura 8 presenta la evolución de los resultados de clasificación de un total de 240 patrones (60 para cada movimiento). Se puede notar que para 5 y 10 repeticiones existe cierta carencia de generalidad, lo que implica en tasas poco robustas debido a la falta de información, especialmente con clasificadores basados en vectores de soporte. Dadas las circunstancias mencionadas para el proceso de entrenamiento, fue limitado a 10 el número mínimo de repeticiones por movimiento. Cabe mencionar que tanto el proceso de entrenamiento como el de evaluación se realizan en-línea, siendo la segmentación de movimiento uno de los principales factores de inclusión de ruido, el cual se incrementa conforme aumenta la incorporación de patrones sin un proceso posterior de depuración, como puede apreciarse en la Figura 8. Los mejores resultados fueron obtenidos utilizando un clasificador Cuadrático Gaussiano y Función de Base Radial.

El método propuesto en este trabajo fue comparado con los resultados obtenidos en Romero et al. (2012b), donde fue alcanzada una tasa de acierto para 3 movimientos de 96.3 %, siendo que en las mismas condiciones el método actual obtuvo un resultado superior. La Tabla 1 muestra la matriz de confusión del método actual aplicado a los movimientos realizados en el estudio anterior (levantar un brazo, levantar dos brazos y gesto de saludo).

Tabla 1: Matriz de confusión en clasificación de los movimientos realizados en Romero et al. (2012b)

	Mov1	Mov2	Mov3	Prec.(%)
Mov1	**60**	1	0	98.40 %
Mov2	0	**59**	0	100.00 %
Mov3	0	0	**60**	100.00 %
Prec.(%)	100.00 %	98.00 %	100.00 %	**99.40 %**

Tabla 2: Matriz de confusión en clasificación de 4 movimientos.

	Mov1	Mov2	Mov3	Mov4	Prec. (%)
Mov1	**59**	0	2	0	96,7 %
Mov2	1	**37**	0	0	97,4 %
Mov3	0	0	**48**	0	100 %
Mov4	0	3	0	**50**	94,3 %
Prec. (%)	98,3 %	92,5 %	96,0 %	100 %	**97,0 %**

La Tabla 2 muestra la matriz de confusión de clasificación de 4 movimientos diferentes, considerando aproximadamente 50 patrones para cada movimiento (Figura 9), los cuales fueron divididos para entrenamiento (50 %), validación (25 %) y test (25 %). Cabe mencionar que el número de patrones de entrenamiento requeridos varía en función del tipo de movimiento, es decir, movimientos similares y de corta duración requieren un mayor entrenamiento; de igual manera, movimientos de mayor duración presentan mejores posibilidades de contener información diferente entre ellos, mejorando la clasificación. Como resultado de experimentos en ambiente interno, 97,0 % de los movimientos fueron correctamente clasificados en experimentos propios y 94,2 % usando 10 movimientos tomados de la base de dados Muhavi-MAS (Singh et al., 2010). El método propuesto realiza reconocimiento en-línea mediante la extracción de patrones tanto en etapas de entrenamiento como en predicción. Este enfoque permite el aprendizaje de diferentes tipos de movimientos (simples y compuestos) mediante repetición.

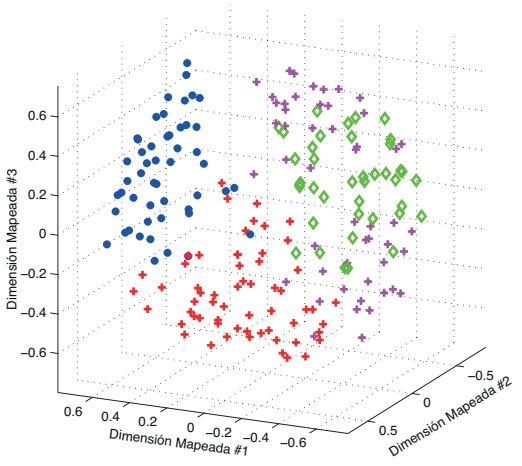

Figura 9: Patrones de cuatro movimientos, mapeados usando kernel t-student.

4. Conclusiones

En este trabajo se presentó una metodología para el reconocimiento en-línea de acciones humanas. El enfoque propuesto hace uso de mapas de profundidad, cuyo uso está siendo cada vez más común en aplicaciones de visión por ordenador, debido al surgimiento de nuevos dispositivos que integran diferentes sensores, con el objetivo de aportar información adicional al proceso de captura. La evolución de este tipo de tecnologías reduciría las limitaciones de los actuales dispositivos y sensores de captura, tanto para enfoques 2D como también 3D. Por otro lado, fue propuesta una forma de aplicar momentos invariantes de Hu como base para el seguimiento y análisis de movimientos humanos, donde fueron seleccionados 3 de los 5 momentos invariantes que proveen información relevante con las siluetas usadas en este trabajo, tomando ventaja de las propiedades de invarianza y cómputo rápido, las cuales posibilitan la extracción en tiempo real de características por imagen. Para el análisis unidimensional de las secuencias de momentos invariantes, la distancia de Mahalanobis dio resultados muy satisfactorios, debido a que permitió obtener información estadística de porciones pequeñas de la señal, aportando a la detección del estado de parada/movimiento, que en este trabajo fue logrado usando sólo 4 frames, permitiendo la detección en-línea de inicio y final de movimiento. Es importante resaltar que el método propuesto de análisis temporal de ventanas con distancia Mahalanobis podría ser aplicado a señales provenientes de otros métodos de extracción de características.

Por ejemplo, si el método es aplicado en identificación de movimientos de pacientes con déficit motor durante fisioterapia, los tremores o posibles oscilaciones de la señal producto de la deficiencia motora pueden ser considerados por la matriz de covarianza, permitiendo la detección de estados de parada y movimiento, siendo un caso similar en el uso de dispositivos inerciales como fuentes de información. Esto último es parte de los futuros trabajos que se están actualmente desarrollando.

Las ventanas dinámicas que acogen las señales después de detectado el inicio de movimiento tienen la ventaja de soportar movimientos complejos o actividades conjuntas específicas, sin necesidad de definir manualmente modelos de transición que limitarían el número de movimientos a reconocer. Esto amplía el alcance del análisis automático no sólo a movimientos básicos como caminar, caer, saltar, etc., sino también al análisis global de ejercicios o actividades realizadas por la persona, posibilitando de esta forma la comparación, medición y caracterización de un grupo de movimientos como una sola actividad.

Como fue mencionado en la sección 2.4, es aplicada la transformada Wavelet Haar para el cómputo posterior de RWE. Las Wavelets Haar son rápidas de calcular, especialmente cuando se trata de señales de corta duración. El tiempo de procesamiento fue de aproximadamente 2 ms para gestos cortos, haciéndolo muy adecuado para aplicaciones de reconocimiento de gestos. Los patrones obtenidos de RWE fueron complementados con el método de SSC, conformando patrones 11-dimensionales para ser clasificados con métodos conocidos como RBF, SVM y K-NN. La extracción de patrones de movimiento se realiza en-línea tanto en etapas de entrenamiento del clasificador como en predicción.

English Summary

Abstract

This paper presents a methodology for online human action recognition on video sequences. It addresses an efficient approach to use invariant moments as image descriptors, applied in processing silhouettes obtained from depth maps. A quick comparison between size-4 windows (equivalent to 4 frames) is performed by computing the Mahalanobis distance, on one of the invariant moment sequences identified as less sensitive to noise and more stable during movement absence. This approach is used for rapid detection of the idle/motion state, which allows the capture of dynamic growth intervals (windows) for further processing, rescuing from the signal contained their temporal and frequential properties. By applying the Haar wavelet transform, three decomposition levels are used for calculating Relative Wavelet Energy (RWE - Relative Wavelet Energy) and SSC (Slope Sign Change), obtaining 11-dimensional patterns. In experiments, 97 % of 4 movements online-captured were recognized correctly, and 10 movements taken from Muhavi-MAS database were recognized with 94.2 % efficiency.

Keywords:

Computer Vision, Depth Maps, Human Action Recognition, Relative Wavelet Energy, Mahalanobis Distance.

Agradecimientos

Este proyecto de investigación es financiado por el Programa Primeros Proyectos, CNPq/FAPES No. 02/2011 y por el CNPq a través de beca de doctorado para el primer autor.

Referencias

Antoniou, A., 2005. Digital Signal Processing. McGraw-Hill.

Broggi, A., Bertozzi, M., Fascioli, A., Sechi, M., 2000. Shape-based pedestrian detection. In: Intelligent Vehicles Symposium, 2000. IV 2000. Proceedings of the IEEE. pp. 215 –220.

Chan, T., Vese, L., feb 2001. Active contours without edges. Image Processing, IEEE Transactions on 10 (2), 266–277.

Chen, Q., Petriu, E., Yang, X., may 2004. A comparative study of fourier descriptors and hu's seven moment invariants for image recognition. In: Electrical and Computer Engineering, 2004. Canadian Conference on. Vol. 1. pp. 103 – 106 Vol.1.

Chockalingam, P., Pradeep, N., Birchfield, S., oct. 2009. Adaptive fragments-based tracking of non-rigid objects using level sets. In: Computer Vision, 2009 IEEE 12th International Conference on. pp. 1530–1537.

Cifuentes, C., Braidot, A., Rodriguez, L., Frisoli, M., Santiago, A., Frizera, A., june 2012. Development of a wearable zigbee sensor system for upper limb rehabilitation robotics. In: Biomedical Robotics and Biomechatronics (BioRob), 2012 4th IEEE RAS EMBS International Conference on. pp. 1989 –1994.
DOI: 10.1109/BioRob.2012.6290926

Franke, U., Joos, A., 2000. Real-time stereo vision for urban traffic scene understanding. In: Intelligent Vehicles Symposium. Proceedings of the IEEE. pp. 273–278.

Garcia-Costa, C., Egea-Lopez, E., Tomas-Gabarron, J., Garcia-Haro, J., Haas, Z., 2011. A stochastic model for chain collisions of vehicles equipped with vehicular communications. Intelligent Transportation Systems, IEEE Transactions on 13, 503–518.

Geronimo, D., Lopez, A., Sappa, D., july 2010. Survey of pedestrian detection for advanced driver assistance systems. Pattern Analysis and Machine Intelligence, IEEE Transactions on 32 (7), 1239–1258.

Gonzalez, R. C., 2010. Digital Image Processing, 2nd Edition. McGraw-Hill.

Grubb, G., Zelinsky, A., Nilsson, L., Rilbe, M., june 2004. 3d vision sensing for improved pedestrian safety. In: Intelligent Vehicles Symposium, IEEE. pp. 19–24.

Haibin, Z., Xu, W., Hong, W., may 2008. Feature selection using relative wavelet energy for brain-computer interface design. In: Bioinformatics and Biomedical Engineering, 2008. ICBBE 2008. The 2nd International Conference on. pp. 1434–1437.

Hu, X., Kong, B., Zheng, F., Wang, S., july 2007. Image recognition based on wavelet invariant moments and wavelet neural networks. In: Information Acquisition, 2007. ICIA '07. International Conference on. pp. 275 –279.

Huang, Z., Leng, J., april 2010. Analysis of hu's moment invariants on image scaling and rotation. In: Computer Engineering and Technology (ICCET), 2010 2nd International Conference on. Vol. 7. pp. V7–476 –V7–480.

Itti, L., Koch, C., Niebur, E., nov 1998. A model of saliency-based visual attention for rapid scene analysis. Pattern Analysis and Machine Intelligence, IEEE Transactions on 20 (11), 1254–1259.

Jones, M., Snow, D., dec. 2008. Pedestrian detection using boosted features over many frames. In: Pattern Recognition, 2008. ICPR 2008. 19th International Conference on. pp. 1–4.

Knoll, P., 2007. Hdr vision for driver assistance. In: Hoefflinger, B. (Ed.), High-Dynamic-Range (HDR) Vision. Vol. 26. Springer, pp. 123–136.

Marsi, S., Impoco, G., Ukovich, A., Ramponi, G., 2007. Video enhancement and dynamic range control of hdr sequences for automotive applications. Advances in Signal Processing (EURASIP) 2007, 9.

Mercimek, M., Gulez, K., Mumcu, T., 2005. Real object recognition using moment invariants. Sadhna - Acad. Proc. Eng. Sci. 30, 765–775.

Miau, F., Papageorgiou, C. S., Itti, L., 2001. Neuromorphic algorithms for computer vision and attention. Proc.Intl Symp. Optical Science and Technology 01 (46), 12–23.

Moeslund, T., Kruger, V., 2006. A survey of advances in vision-based human motion capture and analysis. Computer Vision and Image Understanding 103, 90–126.

Nayar, S., Branzoi, V., 2003. Adaptive dynamic range imaging: optical control of pixel exposures over space and time. In: Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on. pp. 1168 –1175 vol.2.

Park, S., Trivedi, M., 2007. Multi-person interaction and activity analysis: a synergistic track- and body-level analysis framework. Machine Vision and Applications 18, 151–166.

Park, S., Trivedi, M., jul 2008. Understanding human interactions with track and body synergies (tbs) captured from multiple views. Computer Vision and Image Understanding 111 (1), 2–20.

Phinyomark, A., Limsakul, C., Phukpattaranont, P., 2009. A novel feature extraction for robust emg pattern recognition. CoRR abs/0912.3973.

Poppe, R., 2010. A survey on vision-based human action recognition. Image and Vision Computing 28 (6), 976 – 990.

Qiao, Y., Wang, X., Xu, C., june 2011. Learning mahalanobis distance for dtw based online signature verification. In: Information and Automation (ICIA), 2011 IEEE International Conference on. pp. 333 –338.

Rabiner, L., feb 1989. A tutorial on hidden markov models and selected applications in speech recognition. Proceedings of the IEEE 77 (2), 257 –286.

Romero, D., Frizera, A., Bastos, T., jan. 2012a. Movement analysis in learning by repetitive recall. an approach for automatic assistance in physiotherapy. In: Biosignals and Biorobotics Conference (BRC), 2012 ISSNIP. pp. 1–8.

Romero, D., Vintimilla, B., Frizera, A., Bastos, T. F., jun 2012b. Rwe patterns extraction for on-line human action recognition through window-based analysis of invariant moments. In: Robocontrol (2012). Bauru - SP, pp. 20–27.

Rosso, O., Martin, M., Plastino, A., 2003. Brain electrical activity analysis using wavelet-based informational tools (ii): Tsallis non-extensivity and complexity measures. Physica A: Statistical Mechanics and its Applications 320 (0), 497–511.

Rosso, O. A., Blanco, S., Yordanova, J., Kolev, V., Figliola, A., Schurmann, M., Basar, E., 2001. Wavelet entropy: a new tool for analysis of short duration brain electrical signals. Journal of Neuroscience Methods 105 (1), 65–75.

Salas-Lopez, G., Sandoval-Gonzalez, O., Herrera-Aguilar, I., MartÃnez-Sibaja, A., Portillo-Rodriguez, O., Vilchis-Gonzalez, A., 2012. Design and development of a planar robot for upper extremities rehabilitation with visuo-vibrotactile feedback. Procedia Technology 3, 147–156.

Sarvaiya, J. N., 2011. Automatic image registration using mexican hat wavelet, invariant moment, and radon transform. IJACSA - International Journal of Advanced Computer Science and Applications 01 (Special Issue), 75–84.

Singh, S., Velastin, S., Ragheb, H., september 2010. Muhavi: A multicamera human action video dataset for the evaluation of action recognition methods. In: Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on. pp. 48–55.

Soga, M., Kato, T., Ohta, M., Ninomiya, Y., april 2005. Pedestrian detection with stereo vision. In: Data Engineering Workshops. 21st International Conference on. Vol. 01. pp. 20–28.

Viola, P., Jones, M., Snow, D., oct. 2003. Detecting pedestrians using patterns of motion and appearance. In: Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on. Vol. 2. pp. 734–741.

Wang, L., Hu, W., T.Tan, 2003. Recent developments in human motion analysis. Pattern Recognition 36, 585–601.

Yan, L., Casperson, D., Chen, L., june 2011. Survey: Dimension reduction by pattern decomposition. In: Modelling, Identification and Control (ICMIC), Proceedings of 2011 International Conference on. pp. 69 –74.