Un modelo difuso para la estimación de la atención en procesos de interacción robot-persona

Juan Carlos Gámez
Dpto. de Arq. de Computadores, E y TE
Escuela Politécnica Superior
Univ. de Córdoba
14071-Córdoba
jcgamez@uco.es

Antonio González
Dpto. Ciencias de la Comp. e IA
ETS Ingeniería Informática
Univ. de Granada
18071-Granada
A.Gonzalez@decsai.ugr.es

Raúl Pérez
Dpto. Ciencias de la Comp. e IA
ETS Ingeniería Informática
Univ. de Granada
18071-Granada
fgr@decsai.ugr.es

Resumen

Un problema fundamental en la investigación de los procesos de Interacción Robot-Persona es el desarrollo de técnicas que permitan al robot comprender las necesidades de las personas que se encuentran en su entorno. Estas técnicas requieren de procesos en los que se pueda diferenciar a las personas de otros objetos y se pueda comprender, en una primera etapa, si esas personas tienen intención de interaccionar con el robot. Con este objetivo en algunos trabajos previos se propuso un modelo difuso para determinar el grado de interés en la interacción de una persona basado en tres medidas fundamentales: distancia de la persona al robot, ángulo entre el robot y la persona y grado de atención de la persona. Esta última variable está asociada a la orientación de la cara en relación al robot. En este trabajo se propone un método para calcular el grado de atención basado en la selección de características relevantes en la imagen que representa la cara, y en el uso de algoritmos de aprendizaje basados en reglas difusas.

1. Introducción

La robótica desde hace algún tiempo ha ampliado sus posibles campos de aplicación. Ya no se piensa tan solo en un robot trabajando en una cadena de montaje, también están surgiendo robots de servicio, que en muchos casos trabajan en un entorno compartido con los seres humanos. En muchas de estas aplicaciones se requiere que el robot establezca una relación de interacción más natural y próxima con el ser humano. De esta forma, aparecen nuevos retos y problemas que es necesario resolver. Uno de estos problemas es el reconocimiento de las personas, su seguimiento y la detección de los deseos o sus intenciones. En este trabajo nos situamos en la detección del interés que tiene un ser humano en establecer una relación de interacción con el robot. La idea es simple, si tenemos un robot de servicio, y éste está situado en un entorno con diferentes personas, pretendemos que el robot sea capaz de distinguir que persona o personas están interesadas en solicitar un servicio en un momento determinado.

Por supuesto la idea se puede incluir dentro del ámbito de la comunicación entre agentes, donde ahora un agente es físico, el robot, y el otro es un agente humano. Dentro de la comunicación entre agentes hay diversas propuestas. Una primera clasificación distinguiría entre comunicación verbal y no verbal [9]. Un ejemplo de uso de comunicación verbal puede encontrarse en [4]. Por otro lado, respecto a la comunicación no verbal podemos encontrar trabajos que estudian las expresiones faciales [5], las emociones [15] o los gestos [11]. Podemos incluso encontrar trabajos que tratan su estudio desde unos niveles más profundos como el pulso cardíaco, la dilatación de las pu-

pilas o la conductancia de la piel entre otros [8].

En este trabajo nos centramos en el aspecto visual de la comunicación no verbal, realizando un estudio de las imágenes que capta el robot como forma de conocer su entorno, y en concreto de entender las intenciones de las personas.

En un trabajo previo [3] se propuso un modelo visual para la detección del interés. Este modelo está basado en la construcción de un sistema difuso que utiliza información como la distancia entre el robot y la persona, el ángulo que forma la dirección principal del robot y la posición en la que se encuentra la persona y el grado de atención de la misma para poder de forma automática detectar el grado de interés de la persona en la interacción. Las dos primeras variables son relativamente fáciles de calcular, en cambio la tercera, relacionada con la orientación de la cara en relación al robot es bastante más complicada. Con la idea de poder estimar dicha atención se utilizó un algoritmo de aprendizaje que trabajaba sobre las imágenes de las caras, sobre las que previamente se había aplicado un preprocesamiento. Dicho trabajo funcionó de forma relativamente correcta. Sin embargo, tanto el preprocesamiento como el algoritmo de aprendizaje no permitían obtener un conocimiento que se pudiese interpretar. Precisamente, este trabajo se incluye dentro del proceso de estimación de la atención, intentando ahora realizar un preprocesamiento de la imagen que muestre rasgos interpretables, y aplicar un algoritmo de aprendizaje que permita también interpretar el conocimiento que se obtenga.

El trabajo está estructurado como sigue. En la sección 2 se da una breve descripción del modelo básico de estimación de interés, junto con el proceso de detección y seguimiento de las personas. En la sección 3 presentamos el nuevo modelo de detección de la atención, así como la experimentación realizada, finalmente se presentan las conclusiones y trabajos futuros en la última sección.

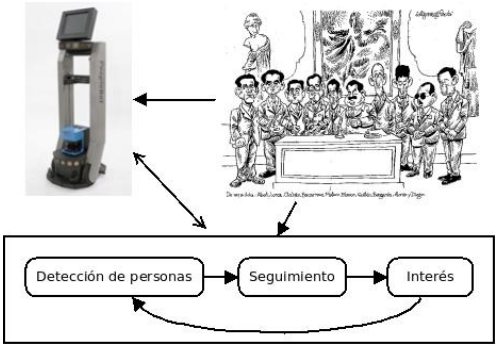


Figura 1: Sistema para interacción robot-persona

2. Modelo de estimación del interés

Uno de los aspectos fundamentales en la interacción persona-robot es el conocimiento que el robot tiene de su entorno, así como de las personas que aparecen en él y cuáles de ellas están interesadas en llevar a cabo una interacción con el robot. Para ello el robot debe ser capaz en primer lugar de identificar a las personas que puedan encontrarse en su entorno. Además debe ser capaz de realizar un seguimiento sobre las mismas para, de esta forma, no perder el contacto con ellas y que la interacción pueda ser fluida. A esto hay que añadir que en estos tipos de relaciones también es importante conocer la disposición que la otra parte tiene en el desarrollo de dicha relación, es por esto que el robot debe ser capaz de conocer el grado de interés en la interacción que tiene la persona que está en su entorno.

En relación a este problema, en [3, 16] se hicieron propuestas en las que se establecen los elementos y fases necesarias para el desarrollo de la interacción. En ellos se desarrolla un sistema modular (ver figura 1) que es capaz de realizar todas las tareas necesarias para conocer, a partir del entorno, las personas que aparecen en el mismo, así como el grado de interés que tienen las mismas en la interacción. Para el desarrollo de estas tareas se utiliza básicamente la información visual captada a través de un sistema de visión.

2.1. Detección de las personas

La primera parte de la interacción consiste en realizar un análisis del entorno en el que se encuentra el robot y a partir de dicho análisis ser capaz de detectar las personas que están en el mismo. Esta tarea se lleva a cabo en el primer módulo del sistema, el módulo de detección de personas, en el cual, para captar el entorno se utiliza un sistema de visión estereoscópica. Para llevar a cabo esta detección se realizan varias tareas. En primer lugar se crea un mapa del entorno que describe el fondo de la escena y está compuesto por los objetos estáticos. Este mapa divide el entorno en celdas de tamaño fijo y se usa para detectar fácilmente los objetos en movimiento en las imágenes siguientes. Posteriormente se crea un mapa de ocupación que junto con el indicará los objetos en movimiento. En este nuevo mapa se buscan los objetos en movimiento susceptibles de ser personas, para lo cual se hace uso de la morfología genérica de las mismas, fijando sobre éstas una zona donde pudiera estar la cabeza. Sobre esas zonas se aplica un detector de caras basado en el detector de caras de Viola and Jones [19] mejorado por Lienhart [13] que está implementado en la librería OpenCV [1].

2.2. Seguimiento

Una vez que se ha detectado la cara, tenemos la certeza de que el objeto se identifica como una persona, y por tanto la siguiente tarea consiste en realizar un seguimiento de la persona durante la posterior secuencia de imágenes. De esta tarea se encarga el módulo de seguimiento que se propone en [17]. Este módulo combina información de posición del objeto con información de color. Por una parte se estima la posición futura de la persona utilizando el filtro de Kalman [14]. Por otra parte se utiliza la información de color, creando un modelo de color del torso de la persona [18]. De esta forma, con la ayuda del color de la ropa de la parte superior del cuerpo y la posición futura obtenida del filtro de Kalman se puede identificar a la persona como aquella sobre la que se está llevando a cabo el seguimiento.

2.3. Estimación del interés

Una vez localizada la persona, la siguiente fase consiste en detectar el interés de la misma en la interacción. En [3] se propone un modelo en el que la detección del interés se realiza mediante la construcción de un sistema difuso que toma como entrada las variable *Distancia*, *Ángulo* y *Atención* y como variable de salida la variable que queremos identificar *Interés*. La variable *Distancia* tiene asociado un dominio compuesto por tres etiquetas lingüísticas *Baja*, *Media* y *Alta* en función de la distancia en metros desde la persona analizada hasta el centro del robot. La variable *Ángulo* hace referencia al ángulo que forma la dirección principal del robot y la posición en la que se encuentra la persona, considerando tan sólo como valores posibles aquellos que están delante del robot o hacia la izquierda o derecha con un ángulo máximo de 90 grados. De esta forma el rango posible de ángulos va de 0 a 180 grados siendo el valor de 90 el ángulo que indica la dirección de la cabeza del robot. La variable *Atención* tiene asociadas también las tres mismas etiquetas pero sobre un rango diferente.

El cálculo de las dos primeras variables es relativamente simple, sin embargo la variable *Atención* es más compleja. Básicamente la idea es determinar la orientación de la cabeza y para ello se pueden seguir distintas aproximaciones.

En un primer trabajo [16] se utilizó una heurística muy simple pero de limitada aplicación en casos prácticos. Dicha heurística consistía en determinar la cantidad de piel que hay en la zona de la imagen que se corresponde con la cabeza. Una vez obtenida la cantidad de piel se establece que si hay una gran cantidad de piel tenemos la cara de frente. Por el contrario, si hay una mínima cantidad de piel la cara está de lado o incluso no está mirando al robot.

Posteriormente [3] se propone un modelo en el que la variable *Atención* se aprende utilizando imágenes que sirven de ejemplo sobre distintas situaciones posibles de la cara. En concreto, se establecen tres clases (A,B,C) que representan tres posibles situaciones de la variable *Atención*, atención Baja, Media y Alta. Se considera la base de datos de ejemplos de

caras, donde previamente se ha asignado una clasificación sobre cada imagen utilizando las clases anteriores. En primer lugar se utiliza un preprocesamiento de las imágenes para reducir la complejidad de las mismas. En concreto, se utiliza el análisis de componentes principales (PCA) [10]. Posteriormente se utiliza el algoritmo de aprendizaje SVM [7] para poder generalizar el conocimiento de dichas imágenes y poder utilizarlo para estimar de esta forma el grado de atención en otras caras no utilizadas en el proceso de aprendizaje.

Finalmente, con la idea de construir las reglas que definen el modelo difuso se utilizó un proceso manual que utiliza el siguiente razonamiento. En primer lugar vemos que la variable más relevante en el modelo es la variable *Atención*. Si esta variable toma el valor bajo o medio, indica que no hay mucho interés en la interacción, por lo que el valor de la variable *Interés* será medio, bajo o muy bajo en función del valor que tomen las otras variables. Esta variable podrá tomar el valor alto o muy alto, sólo cuando la variable *Atención* tome el valor alto. La variable *Atención* toma el valor alto si la cara de la persona que está en la escena está orientada hacia el robot.

3. Estimación de la atención

Como vemos disponemos de una técnica para poder calcular el grado de atención que se requiere como parte del módulo para detectar el interés en la interacción de una determinada persona. Esta técnica funciona correctamente y nos proporciona un modelo general que puede ser de gran utilidad. Sin embargo la utilización de PCA como método de preprocesamiento no permite extraer rasgos interpretables de la imagen, y por tanto no es posible que el conocimiento que pudiese obtenerse por parte de un algoritmo de aprendizaje lo fuese. De todas formas, tampoco SVM permite obtener este tipo de conocimiento.

El objetivo de este trabajo es explorar el uso de algoritmos de aprendizaje basados en reglas difusas en este proceso, y por tanto el definir un preprocesamiento de la imagen que extraiga rasgos interpretables de la misma. A continuación describimos las etapas del proceso.

(a) Baja (b) Media (c) Alta

Figura 2: Muestra de base de datos actual

3.1. La base de datos de caras

Para poder realizar un proceso de aprendizaje de la atención es necesario en primer lugar entrenar al sistema con ejemplos de caras en distintas orientaciones. Así, hemos utilizado una base de datos de caras de uso público para la investigación [2] utilizada fundamentalmente para investigación en psicología. En nuestro caso se han seleccionado 108 imágenes, 18 caras de hombre y 18 de mujer en escala de grises para cada una de las tres orientaciones y con una expresión facial neutra. En la figura 2 mostramos un ejemplo de las imágenes utilizadas y de las tres orientaciones a clasificar (Atención Baja, Media y Alta, respectivamente).

3.2. Puntos Característicos

Nuestro primer objetivo es determinar zonas relevantes de la imagen, para ello utilizamos el concepto de punto característico propuesto en [20]. En concreto utilizaremos la segunda derivada de la gausiana con una elongación 3:1 en la horizontal para detectar franjas de baja intensidad que podrían indicarnos la localización de las cejas, ojos y nariz. Para ello aplicamos en primer lugar un filtro de suavizado gausiano donde los coeficientes de la máscara a aplicar son obtenidos mediante la ecuación 1 donde θ es un valor de ajuste que se aplica para que la suma de la máscara sea 1. El tamaño de la máscara será de $[-h/2, h/2]$ x $[-w/2, w/2]$

donde $h = w = 5 * \sigma^2$. El valor de σ puede ser especificado en tiempo de ejecución y si no es así se toma como 1. Posteriormente, se aplica otro filtro usando la segunda derivada de la gausiana con ratio 3:1 para detectar variaciones bruscas de tonos de gris en horizontal. Para el cálculo de esta máscara se utiliza la ecuación 2 de tamaño $[-h/2, h/2]$ x $[-w/2, w/2]$, siendo en este caso los valores de h y w calculados mediante $w = \frac{2}{5} * anchoImagen$ y $h = w/3$ para conseguir un filtro alargado. Igual que en el caso de la ecuación 1 se utiliza un valor θ como valor de ajuste para que la suma de la máscara sea 1 y el valor de σ puede ser especificado en la ejecución siendo la altura $\alpha = 6 * \sigma^2$. Si no se especifica este valor de σ se calcula como $\sigma = \frac{2}{3} * \text{máx}(h, w)$.

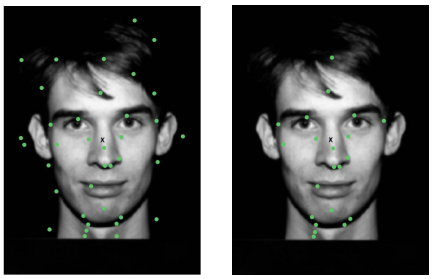
$$coefG(k,l) = \frac{5 * \exp(\frac{-(x^2+y^2)}{2*\sigma^2})}{\theta} \quad (1)$$

$$cLoG(k,l) = \frac{\alpha * \frac{(x^2+y^2-\sigma^2)}{\sigma^4} * \exp(\frac{-(x^2+y^2)}{2*\sigma^2})}{\theta} \quad (2)$$

Finalmente sobre la matriz obtenida con la aplicación de la máscara de la segunda derivada a la imagen suavizada con la gausiana, se realiza una búsqueda de máximos dentro de una vecindad establecida, en nuestro caso 3x3, quedándonos con aquellos que superan un determinado umbral (que nosotros establecemos de forma empírica a 0.01) que serán los puntos característicos de la imagen que utilizaremos para las siguientes fases.

Esta fase nos dará como resultado puntos cercanos a la posible posición de los elementos de la cara. Una vez realizado este proceso y como se puede observar en la figura 3(a) obtenemos demasiados puntos característicos, muchos de ellos irrelevantes para determinar la orientación de la cara. Para reducir el número de puntos característicos no relevantes, aplicamos un detector de bordes tipo Canny, y eliminamos aquellos puntos lejanos a estos bordes. En la figura 3(b) se puede observar el efecto de la reducción.

De forma análoga podemos trabajar sobre la misma imagen pero ahora invirtiendo estos



(a) (b)

Figura 3: Imagen con todos los puntos característicos obtenidos y los cercanos a los bordes

valores. De esta forma, cuando realizamos el cálculo de los puntos característicos, en este caso nos indica la posible localización de los elementos de la cara que pueden ser complementarios a los obtenidos anteriormente (figura 4(b)).

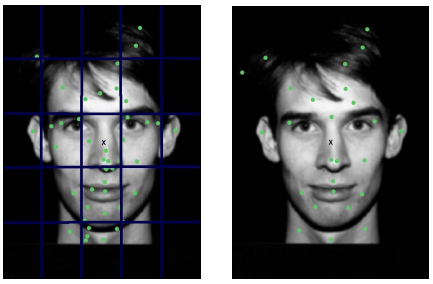
3.3. División en regiones

En la primera fase del procesado, hemos detectado un conjunto de puntos característicos en la imagen. Algunos de esos puntos nos indican la posición de elementos relevantes de la cara. A continuación realizamos una división de la imagen en regiones formando una matriz de tamaño 5x5 como se muestra en la figura 4(a). Con esta división se pretende obtener una agrupación de puntos en las regiones en las que pueden encontrarse los elementos de la cara en función de la orientación de la misma. Adicionalmente a la información sobre los puntos de interés, añadimos nuevas medidas que permitan caracterizar cada una de las regiones de la imagen.

Las medidas que tomamos son las siguiente:

- Media y desviación típica de los valores de los píxeles en cada región: Estas medidas nos indican el tono medio y la dispersión del tono en cada una de la regiones.

- Media y desviación típica del logaritmo de la imagen en cada región: La medias y desviaciones típicas pueden estar afectadas por la iluminación de la imagen. Con



(a) (b)

Figura 4: (a) División en rejillas de la imagen con todos los puntos característicos. (b) Puntos característicos de la imagen invertida

el objetivo de intentar eliminar el efecto de la iluminación nos podemos basar en la formulación simple de imágenes en la que se establece que toda imagen está formada por una componente de iluminación y otra de reflectancia. Si consideramos que una imagen se puede expresar como indica la ecuación 3, podemos reducir el efecto de la iluminación, considerando que ésta es constante y aplicando el logaritmo a la imagen, obteniendo la ecuación 4.

$$img(x,y) = ilum(x,y) * refl(x,y) \quad (3)$$

$$log(img(x,y)) = cte + log(refl(x,y)) \quad (4)$$

3.4. Obtención del modelo

Una vez que hemos fijado las características a tener en cuenta en el modelo, deseamos obtener dicho modelo con la idea de poder clasificar correctamente las imágenes de la base de datos de caras en los tres grados de atención. Para la obtención del modelo utilizaremos el algoritmo NSLV-R [6]. NSLV-R es un algoritmo inductivo de clasificación, que usa como mecanismo de búsqueda un algoritmo genético y que expresa el conocimiento obtenido mediante un conjunto de reglas difusas.

En este caso, tenemos 6 características básicas:

- Proporción de puntos característicos de la imagen en cada región (PI),
- Media de la imagen (MI),
- Desviación típica de la imagen (DTI),
- Proporción de puntos característicos de la imagen invertida en cada región(PIi),
- Media del logaritmo de la imagen (ML) y
- Desviación típica del logaritmo de la imagen (DTL))

medidos en cada una de las 25 regiones diferenciadas en la imagen, con lo que se maneja un total de 150 variables. El valor de cada característica queda identificado con un par de dígitos del 0 al 4, indicando el primero la fila y el segundo la columna de la región. Así, PI23 hacer referencia al número de puntos característicos en la región de fila 2 y columna 3.

Para poder aplicar el algoritmo NSLV-R, es necesario discretizar los dominios de las variables. Para todas las variables se han utilizado dominios con 7 etiquetas lingüísticas *Muy-Baja*, *Baja*, *Media-Baja*, *Media*, *Media-Alta*, *Alta*, *Muy-Alta*. Con objeto de validar el conocimiento obtenido por el algoritmo de aprendizaje, hemos aplicado una validación cruzada sobre 10 particiones obteniendo como resultados: capacidad de clasificación sobre el conjunto de entrenamiento del 99.74 %, capacidad de clasificación sobre el conjunto de prueba del 84.56 %.

En la Tabla 1 se muestra un modelo obtenido por el algoritmo NSLV-R que mostró buen comportamiento en entrenamiento y prueba (99.48 % y 95.45 % respectivamente) compuesto por tres reglas. Podemos observar que es un modelo simple y que incluye algunos marcadores interesantes para estimar la orientación de la cara. Por ejemplo, para la primera regla introduce dos relaciones que son marcadores relevantes para determinar si la atención es alta **(DTI42<≈DT43)** y **(MIL32+MIL22)>≈(MIL12+MIL21)** que involucra comparaciones entre las desviaciones típicas y medias del logaritmo entre regiones adyacentes en la imagen.

Los resultados son satisfactorios obteniéndose un alto grado de predicción en la clasi-

SI PIi10 ES NO Media-Alta PIi14 <= Media MI24 <= Media-Baja DTI14 <= Media DTI23 IS Media DTI42 IS Muy-Baja o Media-Baja (DTI42<≈DT43) (MIL32+MIL22)>≈ (MIL12+MIL21) (MIL21+(MIL12+MIL23)) >= Media-Alta ENTONCES Atención ES Alta (63/0)
SI PI00 <= Media-Alta PI04 <= Media-Alta PI11 ES NO Media PI24 <= Media PI42 <= Media-Baja PIi10 ES NO Muy-Alta PIi11 ES NO Media PIi14 <= Media-Alta PIi43 ES NO Media MI12 >= Media-Baja MI21 ES NO Baja MI33 ES NO Alta DT00 ES NO Baja o Media-Baja DT04 ES NO Bajo o Media-Baja DT14 <= Media-Alta DT22 <= Media-Alta DT42 <= Media DTL12 <= Media-Baja (PI14>≈MIL32) (MIL32+MIL22) ES NO Baja o Media-Baja (MIL21+(MIL12+MIL23)) ES NO Media ENTONCES Atención ES Media (64/0)
SI MI03 ES NOT Media (MI01+(MI02+(MI01+ (DT43*DT41)))) >= Media-Baja (3*MI02+(MI01+(DT43*DT41))) ES Muy-Baja o Muy-Alta ENTONCES Atención ES Baja (64/1)

Tabla 1: Modelo Difuso obtenido para la estimación de la variable Atención

ficación, lo cual indica que los puntos característicos son elementos que aportan información relevante para conocer la orientación de la cara. Sin embargo, se puede observar un cierto grado de sobreaprendizaje. Probablemente este sobreaprendizaje se deba que se están considerando demasiadas variables de entrada y/o que es necesario modificar la estructura de regiones de forma que se adapten mejor a la información que se espera encontrar en cada región, temas que es necesario estudiar con más detalle.

4. Conclusiones y Trabajo futuro

Se ha desarrollado un modelo difuso capaz de medir el interés de una persona en iniciar un proceso de interacción con un robot dentro de un entorno complejo. En concreto, el modelo propuesto se presenta como una alternativa para estimar la variable *Atención*, la variable más relevante para determinar el interés. Las pruebas experimentales muestran que obtiene un conocimiento simple a partir de las características seleccionadas.

Como trabajo futuro se pretende aprovechar mejor la información que aportan los puntos característicos para realizar divisiones más apropiadas de la imagen, detectando las zonas donde potencialmente se encuentran los rasgos faciales (ojos, nariz, boca,...) y determinar a partir de ellos la orientación de la cara.

Agradecimientos

Este trabajo ha sido financiado parcialmente por los proyectos P06-TIC1670, TIN2007-66367 y P09-TIC04813.

Referencias

[1] Web opencv: Open source computer vision library. http://sourceforge.net/projects/ opencvlibrary/, 2010.

[2] Web psychological image collection at stirling (pics). http://pics.psych.stir.ac.uk/, 2010.

[3] E. Aguirre, M. Garcıa-Silvente, R. Paúl, and R. Muñoz-Salinas. A fuzzy system for interest visual detection based on support vector machine. In *ICINCO-RA (1)*, pages 181–190, 2007.

[4] M. Bennewitz, F. Faber, D. Joho, M. Schreiber, and S. Behnke. Multimodal conversation between a humanoid robot and multiple persons. In *Proc. of the Workshop on Modular Construction of Human-like Intelligence at the Twentieth National Conferences on Artificial Intelligence (AAAI)*, 2005.

[5] C. Breazeal. Toward sociable robots. *Robotics and Autonomous Systems*, 42(3-4):167–175, 2003.

[6] Y. Caises, E. Leyva, A. González, and R. Pérez. A genetic learning of fuzzy reltaional rule. In *2010 IEEE World Congress on Computational Intelligence*, page por aparecer, 2010.

[7] N. Cristianini and J. Shawe-Taylor. *An introduction to support Vector Machines: and other kernel-based learning methods.* Cambridge Univ Pr, 2000.

[8] D.K.E.A. Croft. Estimating intent for human-robot interaction. In *IEEE International Conference on Advanced Robotics*, pages 810–815, 2003.

[9] F. Davis. *La comunicación no verbal.* Alianza Editorial, 1998.

[10] G.H. Dunteman. *Principal components analysis.* Sage Publications, Inc, 1989.

[11] D. Kortenkamp, E. Huber, and R.P. Bonasso. Recognizing and interpreting gestures on a mobile robot. In *Proceedings of the National Conference on Artificial Intelligence*, pages 915–921. Citeseer, 1996.

[12] S. Lang, M. Kleinehagenbrock, S. Hohenner, J. Fritsch, G.A. Fink, and G. Sagerer. Providing the basis for human-robot-interaction: A multi-modal attention system for a mobile robot. In *Proceedings of the 5th international conference on Multimodal interfaces*, page 35. ACM, 2003.

[13] R. Lienhart and J. Maydt. An extended set of haar-like features for rapid object detection. In *IEEE ICIP*, volume 1, pages 900–903. Citeseer, 2002.

[14] S.G. Mohinder and P.A. Angus. Kalman filtering theory and practice, 1993.

[15] R.R. Murphy, C.L. Lisetti, R. Tardif, L. Irish, and A. Gage. Emotion-based control of cooperating heterogeneous mobile robots. *IEEE Transactions on Robotics and Automation*, 18(5):744–757, 2002.

[16] R. Muñoz-Salinas, E. Aguirre, M. García-Silvente, and A. González. A fuzzy system for visual detection of interest in human-robot interaction. In *2nd International Conference on Machine Intelligence (ACIDCA-ICMI'2005)*, pages 574–581. Citeseer, 2005.

[17] R. Muñoz-Salinas, E. Aguirre, M. García-Silvente, and A. González. A multiple object tracking approach that combines colour and depth information using a confidence measure. *Pattern Recognition Letters*, 29(10):1504 – 1514, 2008.

[18] D.C.V. Ramesh and P. Meer. Real-time tracking of non-rigid objects using mean shift. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition,*, volume 2, pages 142–149. Citeseer, 2000.

[19] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple. In *Proc. IEEE CVPR 2001*. Citeseer, 2001.

[20] K.C. Yow and R. Cipolla. Towards an automatic human face localization system. In *Proc. 6th British Machine Vision Conference*, volume 2, pages 701–710. Citeseer, 1995.