Uso de técnicas de preprocesamiento de imágenes y aprendizaje para la detección de cambios de atención de una persona en procesos de interacción persona-robot

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

Antonio González, Raúl Pérez, Miguel García-Silvente
{A.Gonzalez,fgr,mgs}@decsai.ugr.es
Dpto. Ciencias de la Computación e Inteligencia Artificial
E.T.S. de Ingenierías Informática y de Telecomunicación
Universidad de Granada
18071-Granada

***Resumen*— Uno de los sensores fundamentales en problemas de interacción persona-robot es la visión. En este trabajo realizamos un estudio en el que, a partir de imágenes de la cara de una persona, extraemos una serie de características relevantes para la clasificación de ciertas actitudes de dicha persona. Gran parte de estas actitudes representan realmente conceptos imprecisos o subjetivos difíciles de definir. En concreto estudiaremos el problema de la detección de la atención que mantiene una persona en relación a un robot de servicio que dispone de una cámara como mecanismo de percepción. Así, nuestro objetivo es utilizar un algoritmo de aprendizaje que, a partir de las características extraídas de la imagen de la cara, permita asignar un incremento o decremento de la atención de la persona en relación a una imagen previa.**

I. INTRODUCCIÓN

La interacción persona-robot es el estudio de la interacción entre las personas y los robots. En estos procesos de interacción la persona se coloca en el eje del modelo de percepción del robot, y sin ninguna duda la visión se convierte en un sensor básico en cualquier procedimiento de interacción.

Hay un gran número de trabajos sobre procesamiento de imágenes que pueden ser de gran utilidad en este tipo de problemas. Podemos encontrar documentos desde los más generales y considerados como "biblias del procesamiento de imágenes" [14], [17], hasta otros trabajos con enfoques más concretos y cercanos a las necesidades de este trabajo [12], donde podemos ver un método basado en bordes para la localización de características de la cara sobre una base de datos de caras, o [23] donde se utiliza el color de los píxeles e información espacial para identificar los distintos componentes de la cara. También tenemos trabajos que se centran en las expresiones faciales [4], otros en emociones [25] e incluso los hay sobre gestos [19], [24], [27].

Por otra parte también encontramos trabajos que tratan la interacción desde el punto de vista del aprendizaje a través de la imitación [26], [29] o incluso el estudio del comportamiento de las personas con los robots [31].

Un problema interesante para la interacción persona-robot es la detección de la atención que mantiene una persona en relación a un robot. Este problema es realmente complejo ya que intervienen factores difíciles de medir y probablemente subjetivos. Obviamente es un problema que requiere de la visión para poder ser resuelto. Hay diversas propuestas que tratan de resolver este problema desde distintas aproximaciones. Por ejemplo, existen trabajos que lo tratan desde el punto de vista de la posición de la cara [30] y otros que versan sobre el reconocimiento de acciones [28].

El estudio de la atención de la persona es un problema que tiene aplicación directa en diversos ámbitos tales como determinar el grado de atención que se muestra ante un sistema de e-learning, o el estudio de los focos de atención que muestran los usuarios ante un sistema de compras web, o la evolución del interés que muestra un espectador ante determinados programas y/o películas de la televisión, con lo que pasaríamos de mediciones de audiencia simplemente como la cantidad de usuarios que ven un determinado canal a conceptos como la calidad del programa o película en función de esta atención.

En este trabajo, para la captación de la atención por parte del robot utilizamos la información recibida por una cámara. Posteriormente realizamos un análisis que nos permite identificar la presencia de una cara en la imagen a partir de ella ciertas características relevantes de la misma. Para la detección de caras podemos encontrar diversos métodos que van desde métodos basados en el conocimiento [35], otros basados en características invariantes como características faciales [21], color de piel [32]; y ajuste por plantillas [20]. Finalmente existen métodos basados en la apariencia entre los que encontramos eigenfaces [33] y basados en filtros de Haar [34] utilizado en nuestro caso para la detección de la cara. Para la extracción de características utilizamos herramientas de procesamiento de imágenes básicas implementadas en OpenCV. Una vez extraídas estas características utilizamos un algoritmo de aprendizaje para obtener el conocimiento requerido en el procedimiento de detección de la atención. Un elemento fundamental en este proceso es la extracción de la información requerida por el algoritmo de aprendizaje. En nuestro caso hemos diseñado un modelo en el que un usuario, observando dos imágenes distintas de la cara de una persona, traslada su impresión sobre si se ha producido un aumento o decremento de la atención de una imagen en relación a la anterior. Este procedimiento permite extraer ejemplos que pueden ser utilizados por el algoritmo de aprendizaje, y se

ha repetido con varios conjuntos de imágenes y por parte de varias personas evaluando estas imágenes.

La estructura del trabajo es la siguiente. En la siguiente sección presentamos los antecedentes de la propuesta. A continuación presentamos las características seleccionadas sobre las imágenes para el proceso de clasificación. En la sección cuarta mostramos el procedimiento para la evaluación del modelo. Finalmente exponemos las conclusiones y trabajos futuros que vamos a desarrollar.

II. ANTECEDENTES

Uno de los problemas interesantes que se presenta en el campo de la interacción persona-robot es detectar cuando un ser humano tiene intención de interactuar. En muchas ocasiones, no es necesario que esta intención sea requerida a través de una orden verbal, sino que el robot la puede deducir de la actitud que adopta la persona en su entorno.

Dentro de la comunicación no verbal [9], encontramos trabajos que proponen sistemas para modelar el interés. En [24] se propone un sistema completo para el cálculo del interés modelado a partir de conjuntos de reglas difusas. El modelo tiene en cuenta tres variables de estado difusas para determinar el interés que se contempla como un concepto difuso. Las variables de estado son *"Distancia"*, *"Ángulo"* y *"Atención"*. *"Distancia"* y *"Ángulo"* representan la posición de la persona dentro de la escena en relación a la posición del robot. *"Atención"* es la variable que valora la actitud de la persona y la más relevante en este sistema para determinar el interés, ya que las reglas que gobiernan el sistema propuesto, requieren que la variable tome un valor alto de *"Atención"* para considerar un valor alto en el interés.

En [24] los autores relacionan el grado de atención con la orientación relativa de la cabeza de la persona con respecto a la posición del robot. Para conocer esta orientación, se hace uso de la información aportada por un detector de piel en el espacio de color HSV [36]. Es la cantidad de piel visible desde la posición del robot la que estima la orientación de la cabeza. Mayor cantidad de piel detectada, indica que la cabeza está orientada hacía el robot.

En [1], se propone un modelo alternativo para estimar la atención usando un algoritmo de aprendizaje. Para ello, se construye un conjunto de entrenamiento formado por imágenes de caras que son etiquetadas en tres clases: "Atención Alta", "Atención Media" y "Atención Baja". Previo al proceso de aprendizaje, se aplica la técnica PCA para extraer las componentes principales más relevantes de cada una de las imágenes. Estas variables son la entrada a un algoritmo de aprendizaje SVM que devuelve un clasificador para estimar la atención.

En [15], se diseña un nuevo sistema difuso para estimar el nivel de atención. Este sistema hace uso de los puntos característicos [37]. En la imagen de una cara, los puntos se encuentran cerca de los rasgos faciales. La propuesta trata de identificar los puntos característicos que forman parte de un rasgo facial. Dicha identificación se realiza mediante la aplicación de un algoritmo de aprendizaje, en este caso del algoritmo NSLV [13], sobre un conjunto de entrenamiento compuesto por trozos de imágenes de caras que se encuentran cerca de un punto característico y que se etiquetan con el rasgo facial al que pertenecen. Para conseguir una mejor aproximación e identificación del rasgo se centran las zonas de la imagen donde se encuentra el punto característico mediante lo máximos de los valores de la zona haciendo uso de las integrales proyectivas [11].

Un problema de todas estas propuestas se encuentra en su utilización en problemas reales con condiciones no controladas. Otro problema importante está relacionado con la dificultad de obtener ejemplos representativos de las diferentes situaciones. La principal diferencia con los trabajos descritos es que el objetivo de esta propuesta se encuentra en la identificación de los cambios de atención y no un valor concreto de la atención, disminuyendo con ello la subjetividad inherente en el proceso de captación de la atención. Además queremos proponer un modelo que sea, en lo posible, robusto y adaptable a diferentes tipos de situaciones.

III. MODELO DE ESTIMACIÓN DEL CAMBIO DE ATENCIÓN

En este trabajo proponemos un nuevo modelo para la estimación de la atención. El sistema, al igual que los descritos en [1], [15], [24] vincula la atención a la orientación de la cara del posible interlocutor con respecto al robot. A diferencia de los anteriores, en este trabajo se propone la identificación del cambio de atención que se produce en la interacción en vez de la identificación de un grado de atención, además de definirse una nueva técnica para la identificación de la cara y los rasgos de la misma.

El proceso general del sistema se puede describir de la siguiente forma:

1. Encontrar en la imagen recibida por la cámara una cara.
2. Encontrada la cara, determinar la posición de los ojos.
3. En función de la distancia entre los ojos, y usando medidas antropométricas, encontrar la posición de la boca.
4. Aplicar un mecanismo de seguimiento para mantener el posicionamiento de los ojos y la boca
   a) Si se detecta que el seguimiento ha perdido el objeto, volver al paso 1.
5. Extraer las distancias y los ángulos que forman ojos y boca, que consideraremos los rasgos relevantes para determinar la orientación de la cara.
6. A partir de estos rasgos, determinar si se ha producido un cambio en la atención.
7. Volver al paso 4.

En las siguientes subsecciones detallamos cada uno de estos pasos.

III-A. Detección de la cara

Para detectar la cara utilizamos los filtros de Haar implementados en OpenCV donde se aplica un detector de caras basado en el detector de Viola and Jones [34] mejorado

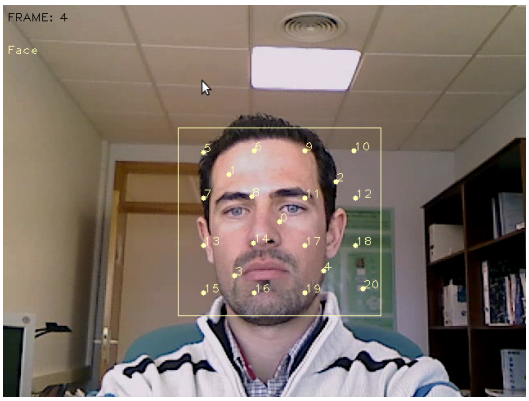
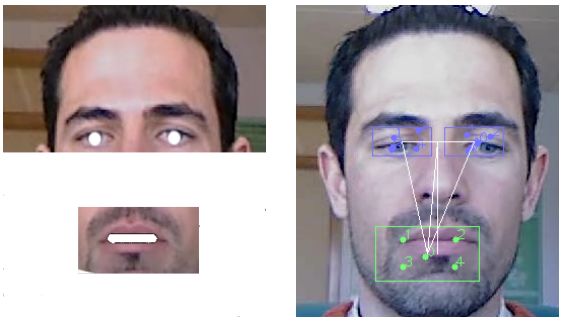


Fig. 1. Detección de la cara



(a) (b)

Fig. 2. Búsqueda de rasgos de la cara

por Lienhart [22]. En [18] se realiza un estudio de esta técnica mostrando su buen comportamiento en comparación con otras técnicas. El método usado requiere que la cara esté situada frontalmente a la cámara. Esto que puede ser una limitación importante en un sistema en general, en el caso de la detección de la atención se puede interpretar como que sólo detectamos la atención de personas que alguna vez mostraron interés en la interacción, y dicha muestra de interés, es que alguna vez miraron al robot.

Tras la utilización de este filtro obtenemos una región de la imagen en la que se encuentra la cara como se muestra en la figura 1.

III-B. Detección de ojos y boca

Para la detección de los ojos en la zona de la imagen donde se detectó la cara se utiliza la transformada de Hough [3] para círculos [16]. Para la detección de la boca usamos la transformada de Hough para líneas [8]. Para hacer más eficiente la aplicación de estos filtros, hacemos uso de la información disponible. En el caso de la detección de los ojos, conocemos que el filtro de Haar detecta una cara frontal, siendo éste bástante robusto frente a cambios de iluminación, color de piel y presencia de algunos complementos en la cara, así como a distancia y posición en la imagen. Los ojos deben encontrarse en la mitad superior y es sobre esa subimagen sobre la que se realiza la búsqueda. Además se debe cumplir que los círculos que se buscan tengan un determinado diámetro en función del tamaño de la cara y deben estar en la misma horizontal (Fig. 2-a), con la posibilidad de que la persona lleve gafas de vista así como otros elementos siempre que el iris del ojo sea visible. Conocida la posición de los ojos y utilizando las medidas antropométricas [7], [10] podemos establecer una subimagen pequeña donde con alta probabilidad se encontrará la boca. Es sobre esta subimagen sobre la que se realiza la búsqueda, igualmente teniendo en cuenta que la boca debe estar presente en la imagen y la línea que se obtiene de la transformada de Hough debe estar centrada respecto a los ojos, horizontal y con un tamaño determinado.

III-C. Seguimiento

En esta etapa se lleva a cabo el seguimiento de los rasgos anteriormente identificados y consecuentemente de la cara. El objetivo del seguimiento es mantener localizada la cara, los ojos y la boca en la imagen sin tener que volver a realizar la detección, y así dotar de un funcionamiento eficiente al sistema.

Para este seguimiento se pueden utilizar diversos métodos como por ejemplo el método de mínimos cuadrados usado en los filtros de Kalman, el método secuencial de Monte Carlo usado en los filtros de partículas (o algoritmo condensation) o el método de flujo óptico de Lucas-Kanade [2] que es el que hemos utilizado en nuestro caso.

En el método de Lucas-Kanade se toman como referencia diferentes puntos de la imagen y se realiza un seguimiento sobre los mismos. En un principio estos puntos son totalmente independientes unos de otros. Además puede darse el caso en el que la diferencia entre la posición de un punto en un *frame* y en el *frame* siguiente sea bastante sustancial, llegando incluso a la pérdida de puntos entre diferentes *frames*. Para evitar esta independencia entre puntos e intentar dar significado al conjunto de los mismos, asociamos al rasgo un conjunto de puntos de seguimiento que lo identifican. De esta forma, cuando se realiza la identificación de cada rasgo, se crean cinco puntos asociados al mismo que están numerados y distribuidos uniformemente como las esquinas y el centro de un rectángulo como se puede ver en la figura 2-b. En esa misma figura se puede observar la creación de un "rectángulo marco" que acotará las posiciones posibles de los puntos de seguimiento para el *frame* posterior. En el caso de que algún punto se obtenga fuera del rectángulo se vuelve a generar su posición, siempre que sea posible, por triangulación con los otros que pertenecen al rasgo. En el caso de que no se pueda regenerar la posición y se perdiesen los puntos volvemos al paso 1 del proceso anteriormente descrito.

III-D. Extracción de atributos a partir de los rasgos

En base a los rasgos detectados, definimos un conjunto de medidas para establecer la orientación de la cara. Estas medidas están formadas por distancias y ángulos entre una serie de puntos definidos sobre la imagen captada de la

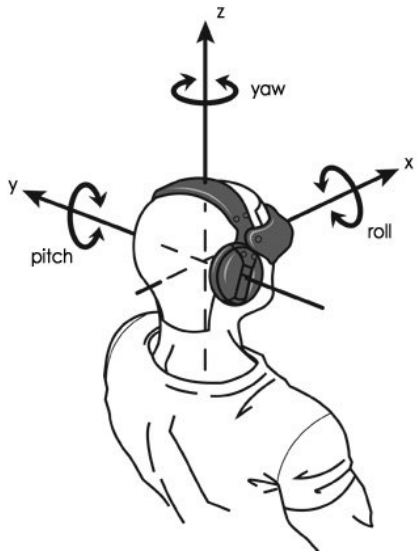


Fig. 3. Ángulos de navegación

cámara. Con estas medidas pretendemos conocer la posición y orientación de la cara en los tres ángulos de navegación: dirección (heading o yaw), elevación (pitch) y ángulo de balanceo (roll) como se muestra en la figura 3.

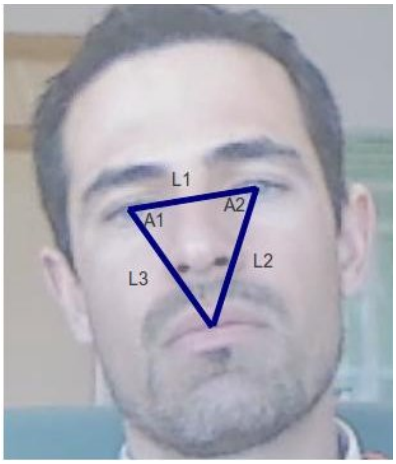
Para la obtención de estas medidas utilizamos:

- **El triángulo formado por los ojos y la boca**: sobre este triángulo tomamos las medidas de los tres lados (**L1**,**L2**,**L3**) y los ángulos (**A1**,**A2**) del mismo como se puede observar en la figura 4(a). Con este triángulo podemos obtenter los distintos valores del ángulo de elevación (pitch) y de dirección (yaw).
- **Los ángulos β y θ**: para la obtención del ángulo de balanceo (roll) y como apoyo al cálculo de los mencionados anteriormente establecemos un eje de referencia en la boca. Sobre ese eje y utilizando el triángulo anterior definimos dos ángulos β y θ como los mostrados en la figura 4(b).
- **Area izquierda (AreaI) y Area derecha (AreaD)**: el eje "Y" del centro de referencia situado en la boca divide al triángulo original en dos triángulos de los que calculamos las áreas que son mostradas también en la figura 4(b).

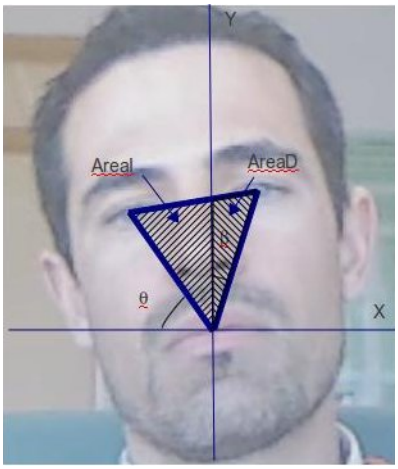
Las medidas comentadas anteriormente son las utilizadas como entrada en el sistema de aprendizaje formando un total de 9 atributos de entrada al sistema.

III-E. Detección de cambio de atención

Esta etapa dentro del proceso se encarga de detectar si se ha producido un cambio de atención en la persona a partir de las medidas propuestas. Para ello, el modelo almacena la última imagen en la que se detectó un cambio en la atención, llamada imagen de referencia, y la compara con la nueva imagen que recibe. Si se determina que en la nueva imagen se aprecia un cambio en la atención en relación a la imagen de referencia, avisa del cambio de atención, y almacena como imagen de referencia la nueva imagen.



(a) Lados y ángulos del triángulo



(b) Ángulos y áreas adicionales

Fig. 4. Variables del sistema difuso

Existen varias posibilidades para establecer esta relación entre cambio de atención y medidas. Una de ellas podría ser tratar de definir una función que contemplando todas las medidas involucradas determinara aritméticamente dicha relación mediante una fórmula. Otra posibilidad, y es en la que estamos interesados, consiste en utilizar un algoritmo de aprendizaje para obtener esta relación. En concreto, usamos el algoritmo NSLV-R [5], [6], una extensión del algoritmo NSLV [13], que es un algoritmo de aprendizaje inductivo que representa el conocimiento mediante un conjunto de reglas difusas, y que usa como mecanismo de búsqueda un algoritmo genético. Una característica útil de NSLV-R para este problema es que puede incluir relaciones entre variables en el antecedente de las reglas. Esta cualidad es interesante en este problema, ya que la detección de cambio se hace a través de la comparación de las medidas obtenidas en la imagen actual con las medidas obtenidas en la imagen de referencia, siendo estas variaciones las que podrán determinar si se producen cambios en la atención de la persona.

En esta propuesta, utilizaremos 5 posibles valores para determinar la variación de la atención entre imágenes, en con-

creto, consideramos: {mucha_más_atención, más_atención, igual_atención, menos_atención, mucha_menos_atención}. La clase *igual_atención* establece que no se produce un cambio significativo en la atención de la persona, mientras que *más_atención* y *mucha_más_atención* muestran un aumento observable de la atención expresado en dos grados, un simple aumento o bien un aumento importante de la atención. Con la misma idea, pero en el sentido opuesto se expresan las clases *menos_atención* y *mucha_menos_atención*.

Una descripción detallada de todo el diseño experimental seguido para la obtención de la base de conocimiento, así como la evaluación del sistema de captación de rasgos y del funcionamiento general del sistema se describen en la siguiente sección.

IV. EVALUACIÓN DEL SISTEMA

En esta sección se describe el diseño de experimentos realizado en este trabajo, tanto para la evaluación del sistema de captación de los rasgos faciales, como para la obtención de la base de conocimiento que regirá el módulo de detección de cambio de atención.

La sección está organizada de la siguiente manera: primero describiremos la forma en la que se han obtenido los datos de partida que servirán para comprobar el comportamiento del modelo de detección de rasgos. A continuación se describe cómo a partir de los datos iniciales y el sistema de detección de rasgos, se define el conjunto de ejemplos de entrenamiento que se usan para obtener el modelo de comportamiento de la estimación del cambio de la atención, incluyendo una breve descripción de las características más importantes del algoritmo NSLV-R que es el usado para extraer la base de conocimiento. Por último, se realiza un análisis de los resultados obtenidos y una evaluación del sistema completo.

IV-A. Obtención de las secuencias de imágenes

Los datos de entrada requeridos son pares de imágenes extraídas de secuencias de imágenes, junto con una etiqueta sobre el cambio de atención que asigna una persona.

En concreto, en esta experimentación se han tomado 11 secuencias de vídeo de 11 personas diferentes. A cada una de las personas que han participado, se les ha sentado delante de una cámara y se le ha indicado que realicen una serie de acciones que implican cambios de su atención en relación a la propia cámara.

La cámara utilizada nos proporciona secuencias de imágenes de 640x480 pixeles, siendo la duración aproximada de cada secuencia de unos 45 segundos.

IV-B. Evaluación del modelo de detección de rasgos

En el proceso de captura de video anteriormente descrito, la primera acción que se le pedía a la persona era que mirara la cámara. Esta primera acción está destinada a realizar la captura de la imagen a partir de la cual se detecta tanto la cara como la posición de los ojos y de la boca. El tiempo medio que emplea el sistema en el reconocimiento completo de la cara y los rasgos, identificando su posición es de 3,32 segundos, de ahí que no necesitemos secuencias de imágenes demasiado largas.

IV-C. Evaluación del modelo de seguimiento de los rasgos

Una vez realizada la identificación de los rasgos con su posición comienza la fase de seguimiento. Durante el tiempo que los individuos están ante la cámara se le indica que lleven a cabo una serie de movimientos con la cabeza para, de esta forma, comprobar el funcionamiento del módulo de seguimiento, al mismo tiempo que obtenemos imágenes que serán utilizadas en la fase de aprendizaje. El número de veces que los puntos de seguimiento de los rasgos de la cara han desaparecido ha sido 4,18 veces de media, lo que obliga al sistema a volver a la fase de detección. La pérdida de los puntos de seguimiento se debe a cuestiones referentes a la calidad de la cámara y al tiempo de preprocesamiento interno de la misma, que provoca la pérdida de *frames* y dificulta considerablemente el seguimiento.

IV-D. Obtención de la base de conocimiento

Para encontrar un modelo basado en el conocimiento que nos permita determinar cuando se produce un cambio de atención haremos uso de un algoritmo de aprendizaje inductivo llamado NSLV-R [5], [6]. Este algoritmo expresa el conocimiento mediante una base de reglas difusas y tiene la capacidad de construir nuevos atributos a partir de los iniciales usando relaciones difusas.

Como todo algoritmo de aprendizaje inductivo, NSLV-R requiere un conjunto de entrenamiento que refleje el comportamiento del sistema que se desea aprender para extraer el conocimiento relevante que lo define.

A continuación describimos la forma en la que se ha obtenido el conjunto de entrenamiento, las características más importantes del algoritmo NSLV-R y los resultados que se han obtenido.

IV-D.1. El conjunto de entrenamiento: Para el conjunto de entrenamiento hemos seleccionado a 11 personas creando un vídeo de cada una de ellas. De cada uno de estos vídeos se han extraído 16 imágenes que, de forma aleatoria, agrupamos en parejas obteniendo 8 parejas por vídeo y un total de 88 parejas de imágenes. Para la clasificación de estas parejas se consulta a 11 expertos mediante la presentación de un subconjunto de 50 parejas diferentes seleccionado de forma aleatoria del total de 88 parejas, para lo cual utilizamos la aplicación mostrada en la figura 5. Con este proceso obtenemos un conjunto de 550 ejemplos de aprendizaje, es decir, pares de imágenes y una etiqueta asociada dentro del conjunto {mucha_más_atención, más_atención, igual_atención, menos_atención, mucha_menos_atención}. En este procedimiento se presentan algunas parejas repetidas a distintos expertos de forma que los aspectos de consenso e incertidumbre sean tratados por el algoritmo de aprendizaje de forma automática. Para obtener un conjunto de aprendizaje mayor hemos asumido simetría en los ejemplos de forma que si al comparar la “imagen 1” con la “imagen 2” obtenemos un cambio de atención etiquetado con *mucha_más_atención*, si la comparamos en sentido inverso, “imagen 2” con “imagen 1” obtendríamos un cambio de atención *mucha_menos_atención*. De esta forma obtenemos

Fig. 5. Recoger la opinión del usuario sobre el cambio de atención

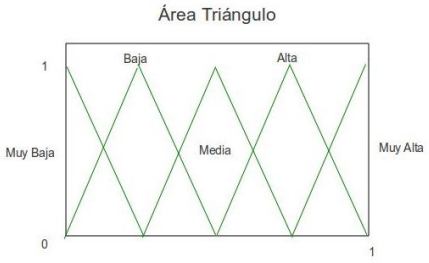


Fig. 6. Dominio difuso de área Triángulo

un total de 1100 ejemplos que utilizamos como conjunto de entrenamiento.

La estructura del conjunto de entrenamiento está formada por 1100 ejemplos en los que encontramos 19 atributos. Los 9 primeros corresponden a los valores de distancias, ángulos y áreas de la imagen de referencia. Los 9 atributos siguientes son los correspondientes a la imagen con la que comparamos. Finalmente aparece el último atributo que indica la clase que corresponde con el cambio de atención de entre los posibles.

IV-D.2. El algoritmo de aprendizaje: NSLV-R [5], [6] es un algoritmo de aprendizaje inductivo que usa como base la estrategia de recubrimiento secuencial, y tiene como elemento principal un mecanismo de búsqueda implementado mediante un algoritmo genético.

NSLV-R es una extensión del algoritmo NSLV [13], del que hereda la mayoría de sus funcionalidades. La diferencia más importante entre ambos algoritmos es el modelo de representación del conocimiento obtenido. NSLV expresa el conocimiento mediante un conjunto de reglas difusas DNF. Una regla DNF tiene la siguiente estructura:

$$\text{SI } X_1 \text{ es } A_1 \text{ y } \ldots \text{ y } X_n \text{ es } A_1$$
$$\text{ENTONCES } Y \text{ es } B$$

donde $X_1, \ldots, X_n$ son las n variables involucradas en el problema de aprendizaje, A_j es un subconjunto de valores del dominio asociado a la variable X_j, Y es la variable consecuente y B es la clase o un valor del dominio asociado a la variable Y.

Por ejemplo, dadas dos variables *Imagen1.AreaI* e *Imagen2.AreaI* que toman valores en el intervalo [0,1] y cuyo dominio asociado está compuesto por cinco etiquetas lingüísticas cuya semántica se muestra en la figura 6, un ejemplo de regla difusa DNF sería:

$$\text{SI } Imagen1.AreaI \text{ es } \{\text{muy baja, baja}\} \text{ y}$$
$$Imagen2.AreaI \text{ es } \{\text{alta, muy alta}\}$$
$$\text{ENTONCES } Clase \text{ es más\_atención}$$

El algoritmo NSLV-R, también expresa el conocimiento mediante reglas difusas DNF, pero a diferencia del anterior, permite incluir relaciones difusas entre las variables de partida. Un ejemplo de este tipo de regla sería el siguiente:

$$\text{SI } Imagen1.AreaI \text{ es } \{\text{media}\} \text{ y}$$
$$Imagen1.AreaI \text{ es aproximadamente igual a } Imagen2.AreaI$$
$$\text{ENTONCES } Clase \text{ es igual\_atención}$$

Para la detección de estas relaciones, NSLV-R toma como entrada una serie de potenciales relaciones entre las variables, y utilizando una medida de información determina cuales son las más relevantes. Estas relaciones entran a formar parte del proceso de aprendizaje.

La posibilidad de considerar relaciones es muy importante en este problema, ya que con el cambio de atención se intenta describir precisamente qué hay en la nueva imagen que es sustancialmente diferente a lo que aparece en la imagen de referencia.

IV-D.3. Resultados experimentales: Como se ha indicado anteriormente, el conjunto de entrenamiento está formado por 1100 ejemplos y 19 atributos: los 9 primeros describen las longitudes, ángulos y áreas definidas para caracterizar la imagen de referencia (la última en la que detectó un cambio en la atención), los 9 siguientes describen las mismas medidas pero de la nueva imagen a etiquetar y la última es la variable de clasificación.

Además, el algoritmo NSLV-R requiere como entrada una serie de relaciones potencialmente útiles para construir nuevos atributos relevantes. Los operadores de relación considerados son los siguientes:

- *menor que* ($<$) y *mayor que* ($>$),
- *aproximadamente igual* ($\approx$),
- *menor o aproximadamente igual que* ($<\approx$).

Además, para cada relación, se indica la lista de las variables a las que se le puede aplicar. En este problema, todas las relaciones consideradas pueden aplicarse a todas las longitudes entre sí, a todos los ángulos entre sí y de igual forma a las áreas entre sí.

En este proceso experimental, hemos seguido el paradigma de la validación cruzada de 10, es decir, el conjunto de entrenamiento inicial se divide en 10 particiones, de las cuales 9 se usan para entrenar el sistema y otra se utiliza para comprobar el grado de clasificación sobre ejemplos no vistos. El proceso se repite 10 veces, de manera que todas las particiones hayan sido utilizadas al menos una vez como conjunto de prueba. Los resultados obtenidos siguiendo este paradigma se muestran en la Tabla I.

Entrenamiento	Prueba	Reglas
96.82 %	94.38 %	33,8

TABLA I
RESULTADOS OBTENIDOS

Los resultados muestran que el algoritmo obtiene una base de conocimiento con muy buena capacidad de predicción. Además, detecta relaciones relevantes, por ejemplo, la relación $Imagen1.AreaI \approx Imagen2.AreaI$ es muy relevante para definir la clase *igual_atención*, o las relaciones $Imagen1.AreaI < Imagen2.AreaI$ y $Imagen1.AreaI > Imagen2.AreaI$ asociadas a las otras clases. Estos hechos nos hacen creer que las medidas tomadas para caracterizar las imágenes son apropiadas.

Por otro lado, el número elevado de reglas denota que existe cierta complejidad para describir el funcionamiento del sistema, y probablemente éste sea uno de los puntos que se requiera mejorar en un futuro.

V. CONCLUSIONES Y TRABAJO FUTURO

V-A. Conclusiones

En este trabajo hemos resuelto la detección del cambio de atención que una persona tiene en el proceso de interacción persona-robot. Para ello hemos considerado que el robot va a captar su entorno mediante una cámara, por lo que hemos utilizado una serie de herramientas de procesamiento de imágenes. Mediante estas herramientas, a partir de la imagen obtenida del entorno, hemos sido capaces de identificar una cara de una persona, así como los ojos y boca de la misma, obteniendo la posición que ocupan en la imagen. Una vez identificadas las posiciones de los ojos y boca, y puesto que esta fase de identificación es relativamente costosa para su funcionamiento en tiempo real, hemos realizado sobre los rasgos identificados un proceso de seguimiento. Hemos utilizado métodos matemáticos para, a partir de estas posiciones, obtener un conjunto de atributos que serán usados como entrada a un algoritmo de aprendizaje supervisado con el que creamos un modelo que representa los cambios en la atención de la persona. Para poder llevar a cabo el aprendizaje hemos creado un modelo de experimentación con el que además realizamos la comprobación del buen funcionamiento del sistema completo.

V-B. Trabajos futuros

Como trabajos futuros se pretende reforzar la detección de los rasgos ampliando la detección a otros rasgos como las cejas. En relacción a las deficiencias detectadas en el módulo de seguimiento, se requerirá un estudio de otras técnicas alternativas más robustas para mejorar este proceso. También será necesario estudiar la razón por la que el problema se ha conseguido modelar con un número tan elevado de reglas, y si es posible simplificarlo. Finalmente, es necesario estudiar el funcionamiento del sistema en un entorno real.

VI. AGRADECIMIENTOS

Este trabajo ha sido financiado por el proyecto de excelencia de la Junta de Andalucía P09-TIC04813.

BIBLIOGRAFÍA

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

[2] S. Baker and I. Matthews. Lucas-kanade 20 years on: A unifying framework. *International Journal of Computer Vision*, 56(3):221–255, 2004.

[3] D.H. Ballard. Generalizing the hough transform to detect arbitrary shapes. *Pattern recognition*, 13(2):111–122, 1981.

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

[5] Y. Caises, A. González, E. Leyva, and R. Pérez. An efficient inductive genetic learning algorithm for fuzzy relational rules. *Por aparcer en: International Journal of Computational Intelligence Systems*, 2011.

[6] Y. Caises, E. Leyva, A. González, and R. Perez. A genetic learning of fuzzy relational rules. In *IEEE International Conference on Fuzzy Systems*, pages 1–8. IEEE, 2010.

[7] G. Chamorro et al. *Manual de antropometría*. Wanceulen Editorial Deportiva, 2005.

[8] O. Chutatape and L. Guo. A modified hough transform for line detection and its performance. *Pattern Recognition*, 32(2):181–192, 1999.

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

[10] L.G. Farkas and I.R. Munro. *Anthropometric facial proportions in medicine*. Charles C. Thomas Publisher, 1987.

[11] Ginés García Mateos. *Procesamiento de caras humanas mediante integrales proyectivas*. PhD thesis, Universidad de Murcia, 2007.

[12] Yulia Gizatdinova and Veikko Surakka. Automatic edge-based localization of facial features from images with complex facial expressions. *Pattern Recognition Letters*, 31(15):2436 – 2446, 2010.

[13] A. González and R. Pérez. Improving the genetic algorithm of slave. *Mathware & Soft Computing*, 16(1):59–70, 2009.

[14] R.C. González and R.E. Woods. *Tratamiento digital de imágenes*. Addison-Wesley Longman, 1996.

[15] J.C. Gámez, A. González, and R. Pérez. Un modelo difuso para la estimación de la atención en procesos de interacción robot-persona. *XI Workshop of Physical Agents 2010*, pages 169–176, 2010.

[16] D. Ioannou, W. Huda, and A.F. Laine. Circle recognition through a 2d hough transform and radius histogramming. *Image and Vision Computing*, 17(1):15–26, 1999.

[17] A.K. Jain. *Fundamentals of digital image processing*. Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 1989.

[18] Vidit Jain and Erik Learned-Miller. Fddb: A benchmark for face detection in unconstrained settings. Technical Report UM-CS-2010-009, University of Massachusetts, Amherst, 2010.

[19] 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.

[20] A. Lanitis, CJ Taylor, and TF Cootes. An automatic face identification system using flexible appearance models. In *British Machine Vision Conference*, volume 1, pages 65–74. Citeseer, 1994.

[21] TK Leung, MC Burl, and P. Perona. Finding faces in cluttered scenes using random labeled graph matching. In *Computer Vision, 1995. Proceedings., Fifth International Conference on*, pages 637–644. IEEE, 1995.

[22] 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.

[23] Christoph Mayer, Matthias Wimmer, and Bernd Radig. Adjusted pixel features for robust facial component classification. *Image and Vision Computing*, 28(5):762 – 771, 2010. Best of Automatic Face and Gesture Recognition 2008.

[24] R. Munoz-Salinas, E. Aguirre, M. Garcıa-Silvente, and A. Gonzalez. 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.

[25] 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.

[26] M.N. Nicolescu and M.J. Mataric. Task learning through imitation and human-robot interaction. *in Models and Mechanisms of Imitation and Social Learning in Robots, Humans and Animals: Behavioural, Social and Communicative Dimensions*, pages 407–424, 2002.

[27] Maja Pantic, Nicu Sebe, Jeffrey F. Cohn, and Thomas S. Huang. Best of automatic face and gesture recognition 2008. *Image and Vision Computing*, 28(5):731 – 731, 2010. Best of Automatic Face and Gesture Recognition 2008.

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

[29] T. Salter, K. Dautenhahn, and R. Boekhorst. Learning about natural human–robot interaction styles. *Robotics and Autonomous Systems*, 54(2):127–134, 2006.

[30] M. Saquib Sarfraz and Olaf Hellwich. Probabilistic learning for fully automatic face recognition across pose. *Image and Vision Computing*, 28(5):744 – 753, 2010. Best of Automatic Face and Gesture Recognition 2008.

[31] M. Scheeff, J. Pinto, K. Rahardja, S. Snibbe, and R. Tow. Experiences with sparky, a social robot. *Socially Inteligent Agents Creating Relationships with Computers and Robots*, page 173, 2002.

[32] S.K. Singh, DS Chauhan, M. Vatsa, and R. Singh. A robust skin color based face detection algorithm. *Tamkang Journal of Science and Engineering*, 6(4):227–234, 2003.

[33] M.A. Turk and A.P. Pentland. Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*, pages 586–591. IEEE, 1991.

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

[35] G. Yang and T.S. Huang. Human face detection in a complex background. *Pattern recognition*, 27(1):53–63, 1994.

[36] M.H. Yang, D.J. Kriegman, and N. Ahuja. Detecting faces in images: A survey. *IEEE Transactions on pattern analysis and machine intelligence*, 24(1):1, 2002.

[37] 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.