Métodos automáticos para el análisis de la expresión oral y gestual en proyectos fin de carrera

Sergio Escalera, Xavier Baró, Petia Radeva
Departamento Matemàtica Aplicada i Anàlisi, Universitat de Barcelona,
Gran Via de les Corts Catalanes 585, 08007 Barcelona
Centro de Visión por Computador,Campus UAB, Edificio O, 08193, Bellaterra, Barcelona
{sergio,xevi,petia}@maia.ub.es

Resumen

La comunicación y expresión oral es una competencia de especial relevancia en el EEES. No obstante, en muchas enseñanzas superiores la puesta en práctica de esta competencia ha sido relegada principalmente a la presentación de proyectos fin de carrera. Dentro de un proyecto de innovación docente, se ha desarrollado una herramienta informática para la extracción de información objetiva para el análisis de la expresión oral y gestual de los alumnos. El objetivo es dar un "feedback" a los estudiantes que les permita mejorar la calidad de sus presentaciones. El prototipo inicial que se presenta en este trabajo permite extraer de forma automática información audio-visual y analizarla mediante técnicas de aprendizaje. El sistema ha sido aplicado a 15 proyectos fin de carrera y 15 exposiciones dentro de una asignatura de cuarto curso. Los resultados obtenidos muestran la viabilidad del sistema para sugerir factores que ayuden tanto en el éxito de la comunicación así como en los criterios de evaluación.

1. Motivación

Con la puesta en marcha de las titulaciones de Grado en el Espacio Europeo en Educación Superior, uno de los objetivos principales es que el alumnado desarrolle una serie de competencias transversales y específicas de cada enseñanza.

La expresión y comunicación oral es una de las competencias más relevantes, considerándose un factor crítico para la vida personal, académica, profesional y cívica de los graduados [3]. En esta dirección, Curtis y Winsor constataron que la comunicación oral era el segundo factor más relevante para la *American Society of Personnel Administrators* [1], realizando posteriormente una encuesta a más de 1000 responsables de recursos humanos, llegando a la conclusión de que una buena capacidad de comunicación oral es importante tanto para la obtención de un puesto de trabajo como para un buen rendimiento en el trabajo [2].

En el caso particular de la Ingeniería Informática, el desarrollo de esta competencia ha estado básicamente relegada a la defensa de los proyectos fin de carrera. El listado y métodos de evaluación de las competencias específicas y transversales de un proyecto fin de estudios ha sido analizado y ampliamente discutido en el ámbito de las ingenierías, donde este tipo de actividades se viene desarrollando desde hace muchos años [5, 6]. En muchos casos, la defensa del proyecto fin de estudios era la primera ocasión en que el alumno se encontraba con la necesidad de comunicar sus resultados de forma oral, sin un entrenamiento previo. En [4] se hizo un estudio sobre el efecto de la aprensión y miedo a la presentación oral sobre la calificación obtenida por los estudiantes. Lo que se deriva de su trabajo es que la aprensión se traduce en peores resultados, y que cuanto más convencidos están los estudiantes sobre sus capacidades comunicativas, más cómodos se sienten y sus calificaciones son mejores. Para poder mejorar la percepción de los estudiantes sobre sus capacidades de comunicación, es necesario generar actividades que requieran comunicar conceptos y/o resultados, generando un buen "feedback"para que puedan ir mejorando sus capacidades.

Con la implantación del Grado en Informática en la Universidad de Barcelona, en algunas asignaturas se han comenzado a realizar pequeñas presentaciones por parte del alumnado para mejorar la comunicación oral y la capacidad de síntesis. No obstante, aún es evidente la necesidad de avanzar en el estudio e implantación de esta competencia.

Como parte de un grupo investigador en Inteligencia Artificial, Visión por Computador, y dentro de un proyecto de innovación docente, se ha desarrollado un sistema automático para el análisis de la expresión oral y gestual de los alumnos. El objetivo inicial

es analizar en que estado se encuentra la capacidad actual de el alumnado a la hora de comunicar ideas, de tal forma que les podamos dar un "feedback"que mejore la calidad de sus presentaciones.

Los métodos automáticos del estado del arte para el análisis del comportamiento humano suelen tener una primera fase de extracción de características y una segunda fase de análisis de los datos extraídos. En relación a la primera fase, muchos trabajos han partido de la extracción de datos mediante el uso de ropas especiales, con sensores o colores específicos que permiten determinar fácilmente la posición y/o aceleración de manos, brazos, cabeza, etc. [7]. Con el objetivo de trabajar en entornos no controlados, otros trabajos se han centrado en la detección de color de piel, movimiento, contornos, o extracción de fondo, lo que permite automatizar y dar más independencia tanto al sistema de reconocimiento como al sujeto que realiza las acciones [8, 9].

El sistema se ha desarrollado para ser utilizado en presentaciones reales, donde los sujetos pueden aparecer tanto con manga corta como con manga larga, sin la necesidad de utilizar ningún elemento artificial para el reconocimiento de las acciones. Se extraen un conjunto de características que nos dan una idea de cómo se está comportando el alumno en la defensa de su trabajo. Además de las características visuales, se han utilizado características básicas de audio para complementar el análisis con el tiempo de habla y pausas, así como combinar rasgos de agitación y comportamiento gestual en situaciones de habla y de no habla.

Una vez extraídas las características que codifican el comportamiento de los sujetos, hacemos uso de clasificadores estadísticos para analizar los datos obtenidos. En particular, en este trabajo se utiliza el método Adaboost [10], que permite aprender clasificadores binarios robustos, a partir de clasificadores binarios simples. Como clasificadores binarios simples se utiliza la combinación de una característica y un valor de corte. Si como medida de evaluación se utiliza la calidad de las presentaciones, este método nos dará las características que mejor separan una presentación buena de una de mala. Además, el algoritmo Adaboost escoge las características por orden de relevancia, lo cual nos permitirá hacer una ordenación y detectar las características más discriminantes en nuestro análisis. En particular, los resultados que hemos obtenido a partir del análisis de 30 grabaciones indican que existe una “correlación” entre la calidad de los proyectos a nivel de contenido y la buena defensa del mismo. Este análisis se ha realizado haciendo uso de las calificaciones finales otorgadas por los profesores así como de las valoraciones asignadas por un grupo de sujetos que visualizaron los vídeos. Además, se obtiene una aproximación de cuáles son las características extraídas que mejor se correlacionan con la opinión de estos observadores.

El resto de este trabajo se organiza de la siguiente manera: el capítulo 2 presenta el sistema desarrollado para la detección y extracción de características. El capítulo 3 realiza la evaluación del sistema a partir de la adquisición de vídeos de presentaciones del alumnado. Finalmente, el capítulo 4 concluye este trabajo.

2. Metodología

En esta sección se describe la parte técnica del sistema para el análisis de la expresión oral y gestual de los alumnos. Los módulos que integran el sistema se muestran en la Figura 1.

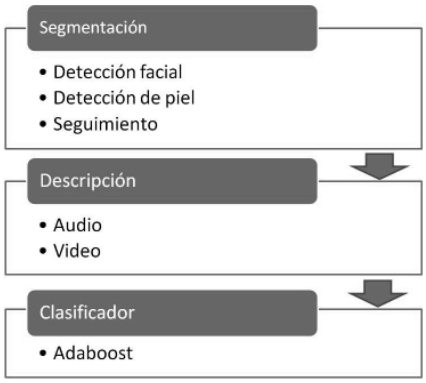

Figura 1: Esquema del sistema de análisis de comunicación oral y gestual.

2.1. Detección de las regiones de interés

El primer paso corresponde a la segmentación de la persona, con el objetivo de aislar las regiones de una imagen que contienen información de interés.

En esta versión del sistema nos centramos en la detección facial y de brazos a nivel de características de vídeo. Para la detección facial, se ha usado uno de los métodos actuales más extendidos, la detección facial de Viola & Jones por cascada de clasificadores [11]. Este método extrae un conjunto de características (Haar-like [11]) de imágenes con caras frontales y se aprende contra un conjunto de características de imágenes sin caras. Este clasificador es entonces probado sobre multitud de regiones de la imagen a diferentes escalas y posiciones. El resultado es la detección de regiones con alta probabilidad de contener una cara.

Además de detectar las regiones faciales mediante el método anterior, los píxeles del interior de la región de la cara nos sirven para determinar con más precisión el color exacto de la piel del sujeto, y de esta forma encontrar las zonas de mayor probabilidad que corresponden a las manos y a los brazos [12]. En la parte superior izquierda de la Figura 2 se muestra la región detectada de la cara y los píxeles de alta probabilidad de pertenecer a la piel de un sujeto.

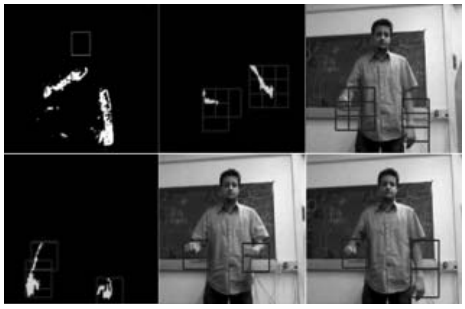

Figura 2: Ejemplo de segmentación en imágenes.

Una vez encontramos los puntos candidatos de pertenecer a una mano o un brazo, el siguiente paso de la segmentación es agruparlos. Para ello definimos las manos y los brazos como las agrupaciones de puntos cercanos que definan una alta densidad de puntos. En la parte izquierda de la Figura 2 se muestran las agrupaciones de mayor densidad por recuadros que corresponden a los brazos. En la parte derecha se muestra la superposición de estos cuadros sobre las imágenes originales.

Una vez tenemos segmentadas las partes de la imagen que queremos analizar, este proceso se repite para todos los frames del vídeo. Teniendo en cuenta que estas regiones se desplazan suavemente en el tiempo, la información de las regiones en frames anteriores se usa para reforzar las detecciones futuras, realizando un proceso robusto de seguimiento de regiones.

2.2. Descripción de las regiones detectadas

Una vez tenemos detectadas las zonas de la cabeza, manos y brazos a través de los métodos de segmentación y seguimiento descritos, hacemos uso de las coordenadas de estas posiciones en el tiempo para extraer un conjunto de descripciones que nos den información a cerca del comportamiento del sujeto.

En el trabajo presentado en [14], los autores definen cuatro indicadores generales que reflejan el éxito de la comunicación y lo evaluan en entornos de interés y dominancia a partir de interacciones sociales. Los cuatro indicadores se definen a continuación:

◇ **Actividad**: Viene definida por la cantidad de habla de un sujeto en un diálogo.

◇ **Estrés**: Corresponde a la agitación corporal de los sujetos en el diálogo.

◇ **Involucración**: Engloba las pautas de conducta que determinan que un sujeto está "sumergido" en el diálogo.

◇ **Copia espejo**: Define la afinidad entre participantes de una conversación a partir de la imitación de gestos y pautas en el habla.

En nuestro caso, el indicador Copia espejo no aparece debido a que sólo hay una persona realizando la presentación. Para el resto de casos hemos definido una serie de descriptores, agrupados por indicador, como se puede observar en la Figura 3 y que se detallan a continuación.

2.2.1. Descriptores de actividad

◇ *Habla*: Porcentaje de tiempo en el que ha estado hablando. Para realizar el cálculo de esta característica se ha utilizado el software de [13], que a partir de un vídeo con audio, obtiene el vector de activación y no activación de la voz en el tiempo.

◇ *No_habla*: Porcentaje de tiempo en el que no ha estado hablando.

◇ *Pausas*: Cantidad de intervalos superiores a dos segundos de duración en los que no ha estado hablando.

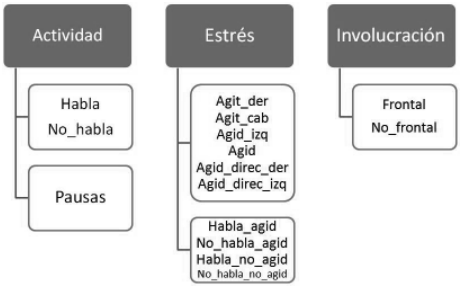

Figura 3: Agrupación de las características.

2.2.2. Descriptores de estrés

◇ *Agit_der*: Promedio de agitación del brazo derecho. Las agitaciones se han calculado a partir de la acumulación de distancias entre las coordenadas de las regiones de frames consecutivos.

◇ *Agit_cab*: Promedio de agitación de la cabeza.

◇ *Agit_izq*: Promedio de agitación del brazo izquierdo.

◇ *Agit*: Promedio de agitación general.

◇ *Agit_direc_der*: Cantidad de desplazamiento realizado hacia la derecha.

◇ *Agit_direc_izq*: Cantidad de desplazamiento realizado hacia la izquierda.

◇ *Habla_agit*: Porcentaje de capturas donde hay alta agitación y se está hablando.

◇ *No_habla_agit*: Porcentaje de capturas donde hay alta agitación y no está hablando.

◇ *Habla_no_agit*: Porcentaje de capturas donde no hay agitación y se está hablando.

◇ *No_habla_no_agit*: Porcentaje de capturas donde no hay agitación y tampoco se está hablando.

2.2.3. Descriptores de involucración

◇ *Frontal*: Porcentaje de capturas frontales (aquellas en las que el sujeto mira al público/tribunal). Aunque a partir del modelo de color podemos realizar el seguimiento facial sin pérdidas, aplicamos el detector de cara frontal [11] para determinar el porcentaje de frames en los cuales el sujeto se dirige al público.

◇ *No_Frontal*: Porcentaje de capturas no frontales (aquellas en las que el sujeto no mira al público/tribunal).

2.3. Clasificación

El objetivo de nuestra herramienta es extraer aquellos patrones que nos diferencien las presentaciones de mejor calidad de aquellas de menor calidad, así como la relevancia de cada una de ellas. Para ello, una vez que se han detectado, seguido, y descrito las características de cada sujeto en cada vídeo que corresponde a una presentación, hacemos uso de clasificadores estadísticos para analizar los datos respecto a la calidad de la presentación. En particular, se ha utilizado Adaboost para el aprendizaje de un clasificador [10].

Utilizando Adaboost se obtiene un clasificador que combina distintas decisiones simples, basadas cada una sobre una única característica. Este método no sólo hace una selección de las hipótesis más relevantes, sino que además proporciona una regla de combinación basada en una suma ponderada de las características.

Detalles sobre este algoritmo se pueden encontrar en [10]. En la parte de evaluación del sistema, este método se usa para encontrar un clasificador que separe entre dos grupos principales de conversaciones, aquellas de mayor “calidad”. Además, también se utiliza para analizar el orden en el cual las características son seleccionadas de mayor a menor relevancia (ranking).

3. Evaluación de la herramienta en vídeos de presentaciones

Antes de presentar los resultados obtenidos, hacemos una breve descripción de los datos analizados, los métodos y los criterios de evaluación utilizados.

◇ *Datos*: Los datos analizados consisten en 15 vídeos filmados en presentaciones de trabajos fin de carrera y 15 en la defensa de proyectos en una asignatura optativa de cuarto curso de Grado en Informática de la Universidad de Barcelona. Todos los vídeos han sido grabados con la misma webcam a una resolución de 640×480 píxels, con un frame rate de 25 imágenes por segundo. Un frame de cada una de las presentaciones se muestra en la Figura 4. Aunque los vídeos son de diferente duración, en el análisis se han considerado 15 minutos para todos los vídeos. Todas las secuencias han sido filmadas en posición frontal al sujeto junto al tribunal, para así poder captar la desviación respecto la posición

frontal y fijación de la mirada del sujeto. Además, todos los alumnos han firmado una hoja de consentimiento de la filmación de sus presentaciones con propósitos de investigación e innovación docente.

◇ *Métodos*: Para el análisis de los vídeos se ha utilizado el sistema descrito en el apartado anterior. Todas las regiones han sido normalizadas respecto el área facial detectada con el objetivo de hacer comparables los valores de las características obtenidas por todos los alumnos. Este paso es importante ya que dependiendo de la distancia del alumno respecto la cámara, el desplazamiento de los píxels puede ser mayor o menor aún cuando la velocidad de agitación entre diferentes sujetos sea la misma. Respecto al clasificador, permitimos que éste haga una selección de las 8 características de más relevancia.

◇ *Evaluación*: Se han realizado dos tipos de evaluaciones. La primera consiste en encontrar aquellas características que mejor correlacionan las notas obtenidas por los alumnos con los patrones de comportamiento. Aunque esta nota final está influenciada por otros aspectos tales como la calidad del trabajo y la escritura de la memoria, queremos analizar si existe una parte comunicativa relevante que influye a las calificaciones finales. En la segunda evaluación se ha realizado una encuesta a 30 sujetos para que visualizaran y evaluaran la presentación de los alumnos. Con estos datos se pretende detectar si existe correlación entre las observaciones de los etiquetadores calificando los vídeos, para posteriormente utilizar el sistema y extraer aquellas características que maximizan la correlación con la previa opinión de los observadores. Ambos experimentos se han realizado considerando problemas binarios, es decir, analizando las características que mejor separan entre dos grupos de presentaciones, que podríamos decir que son las de mayor y menor "calidad".

3.1. Análisis a partir de las calificaciones

En la Figura 5 se muestran ejemplos de las detecciones de las regiones de interés correspondientes a manos y cabeza de algunos sujetos[1].

En este primer experimento se han recuperado las notas asignadas a los alumnos para cada una de las 30 filmaciones. El objetivo es determinar si hay una cierta correlación entre la nota final y la calidad de la presentación. Para ello se han definido dos grupos de 15 vídeos cada uno. Los 15 vídeos del primer grupo corresponden a las filmaciones de las 15 mejores notas, mientras que el segundo grupo corresponde al de las 15 filmaciones con notas inferiores. Después de extraer las regiones y características descritas en las secciones previas, estás se han pasado al clasificador Adaboost para que determine cuáles son aquellas que mejor separan los dos grupos. Para ello se ha lanzado el clasificador varias veces, 8 en concreto. En objetivo es hacer una ordenación de las 8 mejores características. Para ello el clasificador primero determina cuál es la característica de mayor discriminabilidad, y ésta será la primera en la ordenación. Seguidamente, los valores de estas características son extraídas de los datos, y se vuelve a lanzar el clasificador obteniendo la segunda característica de mayor discriminabilidad, y así sucesivamente. En el Cuadro 1 se muestra el orden de las 8 primeras características que mejor separan las presentaciones con mejores notas de las presentaciones con peores notas, y si los valores seleccionados son altos o bajos para discriminar las mejores notas. Se puede observar que la mayoría se centran en la agitación del sujeto, el habla y la mirada frontal para clasificar las mejores presentaciones, mientras que la poca movilidad y paradas en el habla penaliza la presentación. En particular, el clasificador es capaz de separar correctamente las dos particiones de 15 vídeos combinando información de las tres primeras características.

Característica	Valor
Agit_cab	↑
Agit_direc_der	↑
No_Habla_No_Agit	↓
Agit_direc_izq	↑
Agit_izq	↑
Agit_der	↑
Habla	↑
Frontal	↑

Cuadro 1: Ordenación de características seleccionadas por el Adaboost para separar las mejores notas de peores notas. A la derecha se muestra si los valores se seleccionan altos o bajos para discriminar las mejores notas.

[1] Las caras han sido difuminadas de acuerdo al acuerdo de consentimiento firmado por los alumnos

Figura 4: Sujetos de las filmaciones.

Figura 5: Ejemplos de regiones detectadas para diferentes alumnos.

3.2. Análisis a partir de las calificaciones obtenidas por observadores

En este apartado realizamos el mismo análisis que en el caso anterior, pero la separación entre mejores y peores presentaciones se realiza a partir de la opinión de un conjunto de observadores. En particular, las filmaciones se han mostrado a un conjunto de 30 sujetos investigadores y docentes de la Universidad de Barcelona. Obviamente las notas a priori no son comparables ya que cada observador tiene diferentes niveles de rigurosidad en las evaluaciones. Por este motivo, en lugar de una nota numérica, se ha pedido a cada observador que ordenara de mejor a peor cada una de las presentaciones, obteniendo una medida homogénea entre observadores. A partir del orden individual, se ha calculado el orden promedio de cada filmación junto a su varianza. La primera observación interesante es que se pueden diferenciar claramente dos grupos de presentaciones a partir de su ordenación (buenas y malas), a la vez que la varianza de los promedios es reducida, lo cual implica que existe un elevado acuerdo entre las anotaciones realizadas por los observadores. Además, comparando con las agrupaciones realizadas en el apartado anterior, sólo 2 de los 30 vídeos no encajan en la partición definida anteriormente. Esto nos indica que las opiniones de los observadores además están altamente relacionadas con las evaluaciones realizadas por los docentes que pusieron las notas reales de las presentaciones. Con el objetivo de analizar si las características principales cambian por la variación en las agrupaciones producidas por los dos vídeos que cambian de grupo, hemos realizado el mismo análisis que en el apartado anterior. En el Cuadro 2 se muestra el orden de las 8 primeras características que mejor separan las presentaciones con mejores evaluaciones de las presentaciones con peores evaluaciones, y si el clasificador ha seleccionado valores altos o bajos de la característica para

Característica	Valor
Agit_cab	↑
Agit_direc_izq	↑
Agit_der	↓
Frontal	↑
Agit_izq	↑
Habla	↑
No_Habla_No_Agit	↑
No_Frontal	↓

Cuadro 2: Ordenación de características seleccionadas por el Adaboost para separar mejores notas de peores notas en función de las presentaciones y anotaciones de los observadores. A la derecha se muestra si los valores se seleccionan altos o bajos para discriminar las mejores notas.

realizar de forma correcta esta separación. En este caso, aunque el orden en la ordenación ha variado, 7 de las 8 características siguen coincidiendo, dando más relevancia a las características de agitación en las primeras posiciones de la ordenación. En este experimento, el clasificador también es capaz de separar perfectamente las dos particiones de 15 vídeos combinando los valores de las tres primeras características de la ordenación.

3.3. Discusión

Los experimentos realizados muestran la viabilidad del sistema para extraer de forma robusta y automática patrones de comunicación útiles para la expresión oral y gestual de los alumnos.

Aún quedan muchos puntos pendientes de ser analizados. En primer lugar hay algunas situaciones en las cuales la segmentación no es del todo correcta. Algunos ejemplos se muestran en la Figura 6. Básicamente se deben a la unión de las regiones y cambios debidos a la iluminación y oclusiones. El siguiente paso consiste en depurar estas situaciones e incluir nuevos métodos, más robustos que permitan una segmentación y seguimiento con mayor fiabilidad. Además, se podrá incluir información estructural y de expresión facial, tal y como determinar la orientación y estado de las manos y no sólo su localización. El análisis de expresiones faciales también puede permitir añadir nuevas características que enriquezcan la descripción de los sujetos.

También se quiere completar el análisis de audio. En esta versión se detecta el habla y no habla y se combina con información visual. Sería de interés, aunque no se analice el audio a nivel de palabra o nivel semántico, que se pueda además tener en cuenta las variaciones en la monotonía a partir del tono de voz.

Además de mejorar la parte correspondiente a la implementación del sistema, se plantea colaborar con psicólogos y otros especialistas en expresión gestual y verbal con tal de determinar un conjunto más concreto y exacto de características que hagan que el sistema se adapte mejor al diagnóstico de presentaciones, ver como usar exactamente esta información para ayudar a los alumnos, así como pensar en diferentes escenarios donde esta metodología pueda servir también de utilidad para dar un "feedback" u obtener factores de calidad.

Finalmente es importante discutir el coste de implantación y uso de la herramienta. El sistema únicamente requiere de un computador estándar y una cámara digital (que ya suele venir incluida en los dispositivos portátiles). El coste de tal implantación es inferior a los 1000 euros por unidad en la actualidad. Además, el sistema es totalmente automático, lo cual facilita el uso de la herramienta, únicamente teniendo que situar el sistema en el lugar adecuado y haciendo uso de los parámetros especificados en el manual de usuario.

(a) (b) (c)

Figura 6: Ejemplos de detecciones imperfectas. (a) En algunos casos las manos y los brazos interseccionan. (b) En los casos de cambios en iluminación, algunas partes de las regiones a detectar sufren un cambio brusco que los diferencia del modelo de color inicial. (c) Un problema similar ocurre cuando una mano o brazo interseccionа con el opuesto y hay cambios debidos a la iluminación u oclusiones.

4. Conclusión

En este estudio hemos presentado una herramienta para el análisis automático de la comunicación oral y gestual de los alumnos de informática en la defensa de proyectos final de carrera. El sistema es capaz de detectar automáticamente las regiones correspondientes a cara, manos y brazos y extraer un conjunto de características que son analizadas mediante clasificadores estadísticos de Inteligencia Artificial y Aprendizaje Automático. Los resultados obtenidos sobre 30 filmaciones muestran la viabilidad y usabilidad del sistema para obtener valoraciones sobre de la expresión oral y gestual del alumnado, ofreciendo un "feedback" que permita mejorar la calidad de las presentaciones.

El trabajo futuro más inmediato consiste en incrementar la discretización en la clasificación de las presentaciones, incrementando de dos a N categorías de "calidad", para así poder evaluar de forma más precisa la comunicación oral y gestual. También queremos incluir características más precisas para diferenciar agitación y habla entre situaciones de nerviosismo o involucración. Estas situaciones se pueden atacar directamente mediante combinación de características en lugar de indicadores individuales, como por ejemplo: el alumno habla de forma continuada pero se agita sin prestar atención al público, etc.

Referencias

[1] D.B. Curtis, J. L. Winsor, and R.D. Stephens. *National preferences in business and communication education*. Communication Education, Vol. 38 (1), pp. 6-14. 1989.

[2] J. L. Winsor, D.B. Curtis, and R.D. Stephens. *National preferences in business and communication education: A survey update*. Journal of the association of Comunication Administration, Vlo. 3, pp. 170-179. 1997.

[3] T. Allen, *Charting a communicative pathway: Using assessment to guide curriculum development in a re-vitalized general education plan.*. Communicative Education, 51(1) 26-39. 2002.

[4] S. Indra Devi and F. Shahnaz Feroz, *Oral Communication Apprehension and Communicative competence among Electrical Engineering undergraduates in UTeM*. Journal of Human Development and Technology, Vol. 1 Num. 1, June-December 2008.

[5] E. Valderrama, M. Rullán, F. Sánchez, J. Pons, F. Cores, and J.Bisbal, *La evaluación de competencias en los Trabajos Fin de Estudios*, XV JENUI, 2009.

[6] E. Valderrama et al. *Guía para la evaluación de competencias en los trabajos de fin de grado y de máster en las Ingernierías*, AQU Catalunya. 2009

[7] J. Triesch and C. von der Malsburg, *Robotic gesture recognition*, Gesture Workshop, páginas 233-244, 1997.

[8] J. Martin, V. Devin and J. Crowley, *Active hand tracking*, Automatic Face and Gesture Recognition, páginas 573-578, 1998.

[9] F. Chen, C. Fu and C. Huang *Hand gesture recognition using a real-time tracking method and Hidden Markov Models*, Image and Video Computing, volumen 21, número 8, páginas 745-758, 2003.

[10] J. Friedman, T. Hastie and Robert Tibshirani, *Additive Logistic Regression: a Statistical View of Boosting*, Annals of Statistics, volumen 28, páginas 2000-2030, 1998.

[11] P. Viola and M. J. Jones, *Robust Real-Time Face Detection*, Int. J. Comput. Vision, volumen 57, número 2, páginas 137-154, 2004.

[12] M. Jones and J. Rehg, *Statistical color models with application to skin detection*, IJCV, volumen 46, páginas 81-96, 2002.

[13] http://groupmedia.media.mit.edu/data.php

[14] A. Pentland, *Socially aware computation and communication*, Computer, volumen 38, páginas 33-40, 2005.