

udp Escuela de Informática y Telecomunicaciones

UNIVERSIDAD DIEGO PORTALES

ANTE PROYECTO

Reconocimiento de expresiones faciales en imágenes dinámicas utilizando un descriptor basado en rayos de flujo

Autor:
Miguel RODRÍGUEZ

Profesor guía:
Adin RAMIREZ
Comisión:
Javier PEREIRA

3 de abril de 2014

1. Motivación y antecedentes

1.1. Motivación

El reconocimiento de expresiones faciales es parte del reconocimiento de patrones, una rama de la inteligencia artificial encargada de reconocer y clasificar los patrones de comportamiento. Estudios sobre la comunicación oral indican que las expresiones faciales contribuyen con aproximadamente el 55 % de la transmisión del mensaje, lo cual es un valor mucho mayor que la parte verbal (7 %) y la parte vocal (38 %) [1]. Dada la gran importancia de las expresiones en la comunicación es relevante lograr métodos eficientes tanto en la velocidad como en el reconocimiento de estas.

1.2. Antecedentes

En el estudio del reconocimiento de las expresiones faciales existen una variedad de métodos que permiten realizar el reconocimiento de las expresiones canónicas con una alta tasa de aprobación.

1. Expresiones faciales canónicas

Dado que el rostro humano tiene una infinita variedad de posibles expresiones, se establecieron 6 expresiones base o canónicas, las cuales son excluyentes entre sí, además de una expresión neutra. Las cuales son: Surprise (Sorpresa), Happiness (Felicidad), Sadness (Tristeza), Fear (Miedo), Anger (Cólera) y Disgust (Asco).

2. Métodos de reconocimiento de expresiones faciales

El reconocimiento de expresiones faciales se puede dividir en dos grandes aspectos: el primero es la representación vectorial o de características de la cara, y segundo el entrenamiento y clasificaciones con ayuda de algoritmos de clasificación.

a) Representación vectorial o de características de la cara:

Para poder utilizar los algoritmos de clasificación es necesario tener una representación de las características del objeto a clasificar. En nuestro caso para poder etiquetar imágenes o videos de rostros es necesario realizar un proceso de extracción de características.

En el estado del arte existen métodos o algoritmos que permiten la extracción de características, estos métodos se dividen según el tipo de entrada, las cuales pueden ser imágenes estáticas y dinámicas.

1) Métodos sobre imágenes estáticas

Local Binary Patterns (LBP)

Este método consiste en realizar una codificación del píxel de interés con respecto a los valores de sus vecinos. En simples palabras se realiza un proceso de máscara en el cual se restan los píxeles del vecindario con el píxel central, si la resta es negativa o cero se asigna un cero en la posición del vecino, por el contrario si es positiva se asigna un uno. Luego de esto, se realiza una concatenación de los vecinos y se obtiene un código binario que es transformado a base 10. Dicho número es asignado como nuevo valor del píxel visitado.

Luego de realizar el proceso en todos los píxeles se procede a realizar un histograma de los valores el cual es definido como el descriptor de la imagen, siendo este utilizado una para posterior clasificación.

2) Métodos sobre imágenes dinámicas

Volume Local Binary Patterns (VLBP)

Es un método basado en LBP para imágenes dinámicas, el cual se basa en tres variables: L la cantidad de cuadros que entran en la creación del patrón, P el tamaño del vecindario a seleccionar y R el radio del cual se escogen los vecinos. El método VLBP consiste en realizar una codificación basada en LBP, ahora incluyendo la variable temporal L que indica desde

qué cuadro se comienza a realizar la resta del píxel seleccionado con el central. Éste es un proceso muy costoso debido a que mientras mayor sea la cantidad de vecinos a seleccionar, mayor será la cantidad de dígitos binarios, lo cual implica un mayor espectro de números resultantes en la codificación.

Local Binary Patterns - Three Orthogonal Planes (LBP-TOP)

Es un método basado en LBP para imágenes dinámicas, consiste en crear tres planos ortogonales que se intersectan en el píxel de interés, siendo estos el plano XY , XT y YT (cuadro actual, movimiento temporal en X y en Y respectivamente).

Para poder obtener el patrón de la imagen se realiza el mismo proceso de resta con los píxeles del vecindario seleccionado, pero a diferencia de los métodos estáticos que solo se realizan en el plano XY , este también se realiza para los planos XT e YT . Con esto se obtiene un histograma para cada plano, los cuales son concatenados y forman el macro descriptor de la imagen.

b) Entrenamiento y Clasificación

Luego de realizar la extracción de características se procede a realizar la clasificación de estas utilizando algoritmos de clasificación como Support Vector Machines (SVM) y/o Hidden Markov Model (HMM).

Teniendo ya claro como se realiza el reconocimiento de expresiones faciales, también existen técnicas utilizadas en el reconocimiento de patrones que pueden servir para crear nuevos métodos denominados "Híbridos".

1. Bag of Features (BOF) o Bolsa de características.

Esta técnica consiste en crear un universo de todas las características que son extraídas de los casos de entrenamiento, con lo cual se puede tener una representación del espacio vectorial en el cual se está trabajando. Luego se realizan técnicas de clustering para poder crear grupos de características similares. Uno de los grandes problemas de la utilización de esta técnica es la parametrización del número de clusters que se requieren, ya que el valor óptimo de este parámetro se puede encontrar a través de ensayo y error.

2. Optical Flow

Técnica que permite estimar el movimiento de las regiones de interés en las imágenes dinámicas, en general consiste en realizar un seguimiento de la región de interés durante todos los cuadros y obtener un vector que permita representar dicho flujo del movimiento.

2. Descripción de la solución

2.1. Método solución

Para este nuevo método se introduce un nuevo micro descriptor basado en técnicas similares al optical flow, el cual nos permitirá realizar el seguimiento de las regiones de interés o ROI (por sus siglas en inglés) que denominaremos "rayo de flujo" o simplemente "rayo".

La cadena de procedimientos ordenados o "pipeline" de este método se divide en cuatro grandes etapas:

1. Extracción de Micro-Descriptores

Éste proceso se centrará en encontrar las regiones de interés del rostro (ROI), luego para cada una de éstas se realizará un seguimiento con respecto al i cuadro siguiente y se verá cual es el movimiento de éste con respecto al cuadro actual, esto permitirá realizar el cálculo de un rayo de flujo. Al realizar el cálculo de los rayos para cada cuadro se obtendrá un conjunto de rayos que representan la ROI de la imagen.

2. Normalización de rayos y creación del bag of features (BoF)

Luego de la extracción del conjunto de rayos de cada una de las ROI se procede a realizar la normalización de los rayos de tal forma que cada conjunto pueda ser leído en el mismo espacio vectorial.

Luego de normalizar los rayos de cada vídeo, se procede a utilizar la técnica del BoF, dicha técnica tendrá como entrada cada uno de los rayos normalizados obtenidos en el paso anterior. Posteriormente crearemos un número w de clusters que permiten asignar un valor a cada uno de los rayos generados de los vídeos.

3. Generación de macro descriptores

Para poder crear cada uno de estos macro descriptores se procede a realizar un histograma de la imagen resultante del BoF, siendo este el vector de características que será utilizado por el clasificador SVM tanto para entrenar como para realizar las pruebas del sistema.

En el caso de que en la etapa 2 no se logre encontrar una normalización eficiente de los rayos, en este paso se procederá a crear un descriptor por cada uno de los vídeos los cuales serán ingresados al clasificador HMM, el cual permite ingresar los descriptores de forma dinámica.

4. Utilización y prueba de Clasificadores

Al ser un proyecto netamente dedicado a la investigación, se necesitará realizar pruebas con más de un clasificador para poder obtener una gran cantidad de resultados que puedan ser comparados con los métodos propuestos en el estado del arte hasta antes de comenzar la memoria.

Los algoritmos contra los cuales se va a realizar la comparación son los mencionados en los antecedentes.

En esta etapa se procederá a entrenar los métodos propuestos anteriormente, y luego se realizarán las pruebas utilizando técnicas como K -fold cross-validation.

Cabe destacar que esta investigación es de carácter exploratorio, por lo cual, se probarán distintas soluciones a cada una de las etapas anteriormente descritas. Lo cual permitirá probar distintas permutaciones entre cada una de estas soluciones. Al ser un proyecto con un límite determinado de tiempo, hemos decidido que solo se probarán las combinaciones que den mejores resultados, de las cuales la que obtenga el mejor será la seleccionada para el método final de la memoria.

En el caso de no encontrar ninguna posible solución se procederá a describir cada una de las técnicas que se utilizaron, sus resultados y sus respectivas soluciones.

2.2. Objetivos

1. General

Crear un descriptor para el reconocimiento de expresiones faciales en vídeo utilizando micro patrones basados en el seguimiento del flujo de los movimientos del rostro.

2. Específicos

- a) Lograr definir un método de elección de las regiones de interés del rostro.
- b) Lograr definir una codificación que permita la normalización de los rayos de flujo.
- c) Crear una métrica que permita la comparación entre rayos luego del proceso de normalización.
- d) Utilizar y probar distintos algoritmos de clasificación.
- e) Evaluar y comparar el método creado con el estado del arte hasta antes de comenzar la memoria.

3. Metodología de trabajo

Al ser un proyecto de investigación, éste está sujeto a variaciones debido a que se utiliza el método de ensayo y error. La metodología que se utilizará serán dos reuniones semanales para discutir temas relevantes y muestra de avances del proyecto. El proyecto se divide en las siguientes etapas: Diseño del pipeline, Implementación del pipeline, esta etapa implica implementar cada uno de los módulos descritos en la solución, Diseño de casos de estudio, Evaluación y comparación del método, Conclusiones.

4. Cronograma de actividades, hitos y entregables

Fecha	Actividad
24/03/14 - 07/04/14	Análisis del estado del arte.
07/04/14 - 28/04/14	Creación del prototipo.
28/05/14 - 19/05/14	Primera iteración del prototipo.
19/05/14 - 09/06/14	Segunda iteración del prototipo, diseño de pruebas y obtención de resultados.
09/06/14 - 27/06/14	Elaboración del informe.
27/06/14 - 18/07/14	Correcciones y presentación.
11/08/14 - 01/09/14	Tercera iteración del prototipo e implementación de algoritmos del estado del arte.
01/09/14 - 29/09/14	Cuarta iteración del prototipo y comparación con algoritmos implementados.
29/09/14 - 27/10/14	Quinta iteración del prototipo y recolección de resultados.
27/10/14 - 17/11/14	Sexta iteración del prototipo, recolección de resultados y elaboración del informe.
17/11/14 - 28/11/14	Elaboración y entrega del informe.
28/11/14 - 05/12/14	Correcciones y presentación.

5. Resultados esperados

1. Poder crear un método de reconocimiento de expresiones faciales basado en una nueva forma de codificación de patrones, el cual pueda tener una efectividad igual y/o mejor que los métodos del estado del arte.
2. Lograr encontrar un método de normalización de los rayos de flujo que permita la mínima pérdida de la información.
3. Lograr crear una métrica de comparación entre los rayos de flujo que permita la mínima pérdida de la información.
4. Lograr un valor óptimo para el número de clusters necesarios en el BoF.

Referencias

- [1] A. Mehrabian. Communication without words. *Psychology today*, 2:53–56, 1968.
- [2] Caifeng Shan, Shaogang Gong, and Peter W. McOwan. Robust facial expression recognition using local binary patterns. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 2, pages II–370–3, Sept 2005.
- [3] Guoying Zhao and M. Pietikainen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(6):915–928, June 2007.