

Título del tema

Jaime Guzmán

mail@jguzman.cl

Adin Ramirez* y Francisco Claude**

1. Antecedentes y Motivación

1.1. Contexto

Hoy en día es habitual escuchar a las personas comentar sobre las cosas que hace facebook o google en el campo de visión por computador, por ejemplo, facebook reconoce los rostros de las personas y te recomienda etiquetarlos o te dice “esta persona puede ser conocido tuyo” por medio de una fotografía, ó como “google image” que con tan solo escribir una palabra este te entrega un conjunto de imágenes extraídas de la web relacionadas con la palabra escrita.

Dados estos ejemplos y junto con la posibilidad de investigar nuevas tecnologías relacionadas con esta área de la computación, y tomando en cuenta el estado del arte en el que está este tipo de investigaciones, presentaremos una manera eficiente, ligera y rápida para la identificación de objetos y expresiones faciales.

Nos centraremos en la investigación de un nuevo mecanismo de descripción de características con el fin de crear una estructura que represente eficazmente dichas características y sea ligero para su procesamiento en el reconocimiento de objetos y expresiones.

1.2. Trabajos relacionados

En esta área se encuentran variados trabajos que nos dan una idea de lo que se ha descubierto y propuesto hasta el momento. Trabajos tales como los presentados por Ramanan et al. [?], Ramirez et al. [?], Bay et al. [?], y Lowe [?], nos servirán de guía para realizar nuestra investigación. A continuación se detalla en rasgos generales cada uno de ellos.

- Ramanan et al. [?] presenta un modelo basado en partes, entendiendo por modelo, una abstracción teórica del mundo real, el cual permite hacer predicciones sobre la información que estamos analizando y por modelo basado en partes, a la utilización de varias partes de la imagen con el fin de determinar si existe un objeto de interés en ella. En este trabajo se busca optimizar *Dalal-Triggs detector* el que utiliza un único filtro de HOG [?]. Este tiene como principal objetivo obtener características específicas del objeto mediante la dirección de los gradientes o bordes, esto es implementado de la siguiente forma: Se divide la imagen en pequeñas porciones, llamadas celdas, las cuales generan otros histogramas de gradientes o bordes por cada pixel dentro de esta celda, combinando estos se obtiene la representación del descriptor. La mejora realizada sobre este método es que en paralelo se computa una función que maximiza los puntos obtenidos por HOG, lo cual permite capturar más características en el mismo espectro de muestreo.
- Ramirez et al. [?] presenta un nuevo descriptor de características locales, llamado Local Directional Number Pattern (LDN), para reconocimiento de rostro y expresiones. Este estudio comenta que la eficiencia de un descriptor depende de su representación y la facilidad de extracción del rostro. Idealmente un descriptor debería tener una alta varianza con clases muy distintas (diferentes personas o expresiones), por otra parte si las clases son semejantes (misma persona o expresión), su varianza debería ser inferior o cero.

Comúnmente existen dos enfoques para la extracción de características en el rostro: basados en características geométricas y basados en apariencia, nos enfocaremos en este último ya que es el que más nos puede servir para a lo que deseamos hacer por su naturaleza, puesto que este método utiliza filtros que crean características generales o en una área del rostro en específico para crear características locales, con el fin de extraer los cambios de apariencia en la imagen del rostro.

*Profesor guía

**Profesor comisión

- Lowe [?] presenta un método para la extracción de características invariantes de una imagen que son utilizadas para hacer una comparación entre diferentes vistas o escenas de un objeto. Las características son invariantes a la escala de la imagen y la rotación que esta pueda presentar. Por el lado de los descriptores, estos se representan con la dirección y el gradiente que poseen los puntos de interés encontrados en etapas previas, luego se preprocesan los datos con una función de peso gaussiana, la cual asigna un peso a cada uno de los puntos. El propósito de este filtro es evitar que las muestras sufran variaciones.
- Bay et al. [?] presenta un detector y descriptor que es invariante a las rotaciones, llamado SURF (Speeded-Up Robust Feature). Como lo que nos interesa son los descriptores, SURF utiliza un descriptor basado en distribuciones. Para SURF la creación de un descriptor debe ser único e invariante al ruido. Para esto cada descriptor describe una distribución de intensidad de los puntos adyacentes al punto de interés, similar a lo que se postula en Lowe [?], pero con la diferencia que no se utiliza el gradiente sino que una distribución de primer orden basado en los wavelet the Haar.

Con estos antecedentes previos podemos comenzar a preparar nuestra hipótesis para utilizar descriptores en imágenes temporales (vídeos), con el objetivo ya dicho antes de reconocer de mejor forma objetos y expresiones de personas dentro de vídeos o ráfagas de imágenes.

1.3. Motivación

Nuestra principal motivación es poder generar un descriptor eficiente, en tiempo de ejecución y que sea ligero con el fin de poder ser utilizado, por ejemplo, en aplicaciones móviles. En detalle, dicho descriptor podrá ser utilizado para crear aplicaciones que nos permitan, por ejemplo, pasear por un museo y dentro de éste, visualizar un elemento; y mientras lo observamos obtenemos información sobre este mismo de forma rápida y precisa.

Incluso poder ir por la calle, ver a personas y poder saber quienes son antes de que nos saluden, detectar objetos interesantes y obtener información sobre ellos, llevar la computación más allá del viejo y olvidado ordenador.

2. Descripción de la solución

La idea principal es generar un descriptor basado en partes utilizando métodos de reconocimiento de patrones temporales.

- Objetivos generales.
 - Combinar los diferentes métodos para generar un nuevo descriptor.
 - Analizar técnicas basadas en partes y basadas en apariencia, tanto para objetos como para expresiones faciales en secuencia de imágenes.

■ Objetivos específicos.

- Analizar y evaluar el comportamiento del descriptor con conjuntos de imágenes, para poder obtener métricas que nos permitan identificar que tan buena es nuestra propuesta para identificar objetos.
- Analizar y evaluar el comportamiento del descriptor con diversas imágenes, para poder obtener métricas que nos permitan identificar que tan buena es nuestra propuesta con expresiones faciales.
- Analizar la precisión del descriptor, es decir, obtener métricas que nos permitan verificar mediante experimentos que tan eficiente es el algoritmo propuesto.
- Analizar el rendimiento y velocidad de computo de nuestra propuesta con métricas específicas para analizar dicho comportamiento.

3. Metodología de trabajo

3.1. Etapa 1

Se necesita generar un marco de trabajo donde se pueda probar el algoritmo diseñado por nosotros para luego hacer las pruebas respectivas junto con los demás descriptores y modelos que aparecen en el estado del arte.

3.2. Etapa 2

Luego con este marco de trabajo realizaremos el pre-procesamiento de la información, para así poder probar nuestra propuesta con el fin de poder clasificar dicha información como corresponde.

3.3. Etapa 3

Etapa donde revisaremos si nuestra propuesta es viable o no, con el fin de asegurar el trabajo posterior.

3.4. Etapa 4

En esta etapa validaremos nuestro modelo, comparándolo con los modelos ya existentes en el estado del arte, tomando en cuenta diversas características con el fin de generar un análisis más riguroso de nuestra propuesta.

4. Cronograma de actividades, hitos y entregables

Fecha	Actividad
21/03/2014	Presentación anteproyecto (firmado por profesor guía y comisión).
02/04/2014	Entrega resultados anteproyectos.
04/04/2014	Entrega anteproyectos corregidos.
09/04/2014	Entrega resultados correcciones.
25/04/2014	Marco de trabajo.
23/05/2014	Primer prototipo de propuesta.
13/06/2014	Resultados parciales.
27/06/2014	Entrega Memoria Titulo 1 firmada por profesor guía.
11/07/2014	Fecha límite para que la comisión entregue correcciones.
18/07/2014	Fecha límite para que se realicen correcciones.
20/08/2014	Avance.
10/09/2014	Avance.
08/10/2014	Entrega Descriptor.

5. Resultados esperados

Esperamos poder crear un nuevo descriptor robusto con ayuda de los trabajos hechos con anterioridad relacionados con este tema para poder identificar tanto objetos como expresiones. Diseñaremos nuestros algoritmos con el fin de generar un descriptor eficiente (con la idea de que pueda ser utilizado en un sistema móvil), para así ayudar a las personas a obtener información del mundo real de forma más fácil y dinámica.

A su vez queremos analizar y verificar que nuestra propuesta cumple con los estándares vistos en el estado del arte de nuestra investigación, para así proponer nuestro método como, un método robusto y eficiente en comparación con los que existen en la actualidad.