

Seguimiento de objetos en secuencias de imágenes RGB-D

Tesis de licenciatura

Mariano Bianchi

Facultad de Ciencias Exactas y Naturales

Miércoles 18 de Marzo de 2015

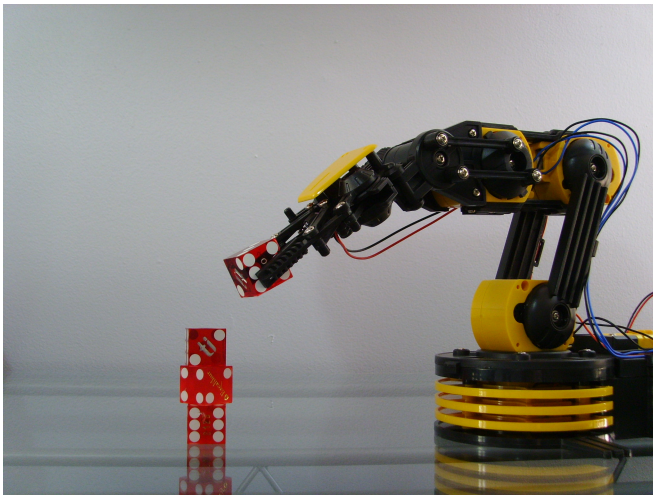
Si nos organizamos...

- 1 Introducción
 - Aplicaciones
 - Vamos por partes... del título
 - Objetivos
- 2 Sistema de seguimiento
 - Sistema RGB
 - Sistema en profundidad
 - Sistema RGB-D
- 3 Resultados
 - Base de datos
 - Selección de objetos para las pruebas
 - Métricas
 - Análisis
- 4 Conclusiones y trabajo a futuro

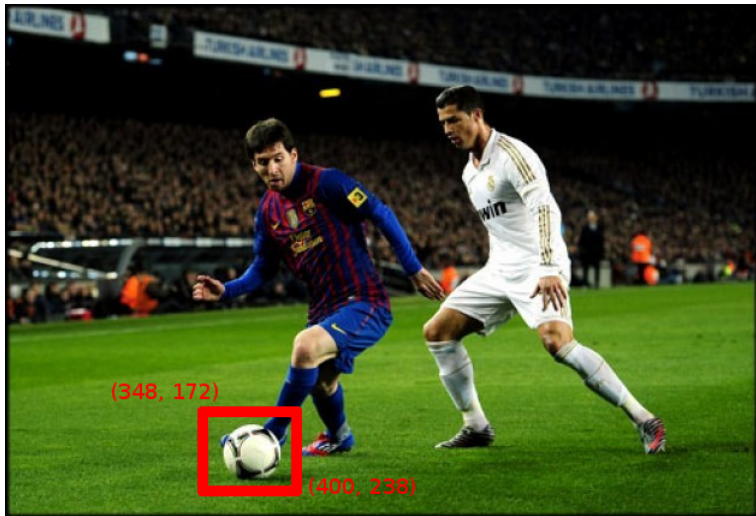
Aplicaciones



Aplicaciones



Seguimiento de objetos en secuencias de imágenes RGB-D



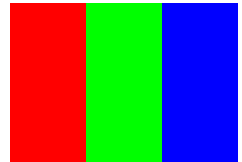
Seguimiento de **objetos** en secuencias de imágenes RGB-D



Seguimiento de objetos en secuencias de imágenes RGB-D



Seguimiento de objetos en secuencias de imágenes **RGB-D**



Seguimiento de objetos en secuencias de imágenes RGB-D



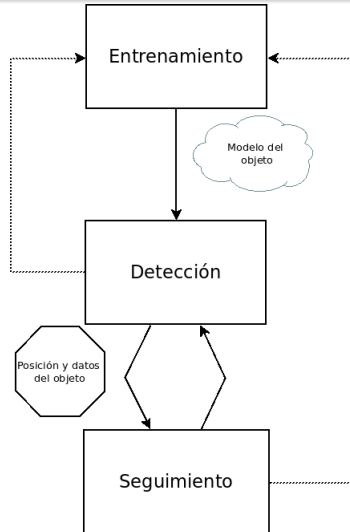
Ambos respetan la patente de PrimeSense  PrimeSense



Objetivos

- Implementamos, estudiamos y evaluamos un sistema de seguimiento RGB-D, enfocándonos en la etapa de seguimiento
- Comparamos métodos de seguimiento en RGB y en profundidad y analizamos sus ventajas y desventajas
- Reportamos resultados para un sistema de seguimiento RGB-D completo, no encontrados en la literatura

Sistema de seguimiento



Distintos esquemas para el mismo problema????

- Sistema RGB completo y funcional
- Sistema en profundidad completo y funcional
- Combinarlos de la mejor manera posible
- Obtener resultados comparables frente a métodos existentes

Detección RGB

En esta etapa usamos un método llamado *template matching*.
Para utilizarlo necesitamos:

- *templates*: Una o más imágenes del objeto a detectar tomadas de diferentes ángulos
- *escena*: Un frame de un video o imagen en donde se desea ubicar el objeto
- Opcionalmente podemos utilizar máscaras que segmenten al objeto en cada template.

Detección RGB

PONER ESTAS IMAGENES

- 1 Un template entero
- 2 Un template segmentado
- 3 Una escena

Detección RGB

Pasos del algoritmo de template matching

Para cada template proveniente del entrenamiento y para cada pixel de la escena, seguir estos pasos:

- 1 Tomar un rectángulo de la escena del tamaño del template cuya esquina superior izquierda sea el pixel actual
- 2 Compararlo con el template (ejemplo: diferencia cuadrática pixel por pixel)
- 3 Si la comparación está por debajo de un umbral predefinido y es el mejor valor encontrado, guardar la ubicación del pixel

Una vez recorrida toda la imagen, se devuelve la ubicación del “mejor recuadro”. Si no se encontró ninguno por debajo del umbral, se indica que no se encontró el objeto en la imagen.

Entrenamiento RGB

Como el algoritmo de detección elegido es *template matching*, esta etapa consta simplemente de obtener distintos templates del objeto que se desea seguir, tratando de cubrir las distintas “caras” del objeto desde varias alturas. Por ejemplo:

PONER UN PAR DE TEMPLATES

Seguimiento RGB

Tomando la ubicación del objeto en el frame anterior, se explora en un área cercana (mucho menor que el área de la escena) de manera similar a la de *template matching* en búsqueda del objeto deseado. Para cada recuadro explorado se siguen estos pasos:

- 1 Calcular histograma del recuadro
- 2 Calcular histograma del recuadro del frame anterior
- 3 Calcular histograma del último template “matcheado”
- 4 Comparar los histogramas 1-2 y 1-3
- 5 Si ambas comparaciones están por debajo de sus respectivos umbrales y son las mejores, se guarda la ubicación del recuadro (esq. superior izquierda)

Si se guardó una ubicación, se la devuelve como resultado. Si no, se indica que no se encontró el objeto y se vuelve a la etapa de detección

Detección en profundidad

Para esta etapa vamos a usar dos métodos de alineación de nubes de puntos y combinarlos. Para usarlos necesitamos:

- *vista del objeto*: Una nube de puntos del objeto parcial o completa (similar al concepto de template en RGB)
- *escena*: Una nube de puntos de la escena donde se desea encontrar al objeto

Detección en profundidad - Método 1

El método se llama Iterative Closest Point (ICP)

Algoritmo ICP

Dadas dos nubes de puntos A y B, se siguen los siguientes pasos hasta que alguna condición de corte se cumpla:

- 1 Se calculan los “puntos más cercanos” entre A y B
- 2 Se calcula una transformación que una los puntos cercanos de A en B
- 3 Se aplica la transformación sobre A

Al terminar se devuelve la transformación T proveniente de aplicar todas las transformaciones parciales junto con el error cuadrático medio entre los puntos más cercanos de $T(A)$ y B

Detección en profundidad - Método 2

El método usado es una variante de un esquema RANSAC

RANSAC

Dadas dos nubes de puntos A y B, se calculan sus descriptores (features) y se siguen estos pasos:

- 1 Tomar N descriptores random de A
- 2 Encontrar sus correspondencias en B
- 3 Descartar si los poligonos resultantes no son isométricos
- 4 Si lo son, estimar una transformación entre los descriptores
- 5 Transformar y calcular error cuarático medio
- 6 Almacenar transformación si tiene el menor error hasta ahora

Al terminar se devuelve la mejor transformación T encontrada y el error cuadrático medio entre los puntos más cercanos de $T(A)$ y B

Detección en profundidad - Combinación

El método final comienza con una nube de puntos del objeto O y otra de la escena S . Se segmenta la escena de manera similar a *template matching* y para cada segmento se siguen estos pasos:

- Aplicar RANSAC entre O y S
- Si no es exitoso, se continua con el siguiente segmento
- Si fue exitoso, aplicar ICP entre RANSAC(O) y S
- Si ICP fue exitoso, se almacena el resultado y se sigue con otro segmento

Si se almacenó algún segmento, se filtran los puntos de la escena correspondientes al objeto en ese segmento, se obtiene su ubicación en coordenadas RGB y se pasa al siguiente frame

Entrenamiento en profundidad

Como el algoritmo de detección elegido es *template matching*, esta etapa consta simplemente de obtener distintos templates del objeto que se desea seguir, tratando de cubrir las distintas “caras” del objeto desde varias alturas. Por ejemplo:

PONER UN PAR DE TEMPLATES

Seguimiento en profundidad

Tomando la nube de puntos del objeto segmentado en el frame anterior y la nube de puntos de la escena en el frame actual, se siguen estos pasos:

- 1 Se filtran los puntos de la escena cercanos a la ubicación del objeto en el frame anterior
- 2 Se aplica ICP entre el objeto y los puntos filtrados de la escena

Si ICP fue exitoso, se devuelve la posición del objeto y se continua en el siguiente frame. Si falla, se vuelve a la etapa de detección sobre el mismo frame.

Sistema RGB-D

Entrenamiento

Se unen los entrenamientos de RGB y profundidad

Detección

Se utiliza primero la detección RGB y si es exitoso se transforman estos datos para utilizarlos con la detección en profundidad. Una detección RGB-D es exitosa siempre que lo sean ambas detecciones por separado.

Seguimiento

Primero se aplica el seguimiento en profundidad. Si es exitoso se intenta mejorar con el seguimiento en RGB, partiendo del resultado en profundidad. Un seguimiento RGB-D es exitoso si el seguimiento en profundidad es exitoso. La parte RGB se usa solo

Base de datos

- 300 objetos de uso diario
- 51 categorías
- Segmentados en RGB y en profundidad



- 8 escenas
- Distintas instancias y categorías de objetos por escena
- Anotadas en RGB

Selección de objetos para las pruebas

Para la selección de parámetros

- 3 objetos: taza blanca, gorra roja, bowl verde-agua
- 2 escenas: escritorio 1 y 2

De control

- 3 objetos: taza roja con dibujos, lata, caja de cereales
- 2 escenas: mesa grande y mesa pequeña

Comparación de resultados vs. ground truth

IMAGEN DEL PDF DE PASCALLIN VOC METRIC QUE
MUESTRA LO DE INTERSECCION / UNION

Selección del umbral de solapamiento

FORMULA DEL ACCURACY?

GRAFICO DE UMBRAL VS ACCURACY

Comparando seguimientos

RGB	Obj	% seg.	% solap.
	Taza 1	47.89	34.48
	Gorra	96.97	55.68
	Bowl	6.19	12.58
	Taza 2	35.9	29.54
	Lata	0	0.01
	Caja	67.62	52.11

Profundidad	Obj	% seg.	% solap.
	Taza 1	92.16	66.29
	Gorra	92.67	56.85
	Bowl	63.83	40.60
	Taza 2	80.51	55.79
	Lata	10.63	12.34
	Caja	30.62	44.94

RGB-D	Obj	% seg.	% solap.
	Taza 1	90.05	65.07
	Gorra	90.82	42.83
	Bowl	68.54	42.88
	Taza 2	48.33	35.98
	Lata	23.67	18.81
	Caja	23.29	35.03

Comparando seguimientos

RGB	Obj	% seg.	% solap.
	Taza 1	47.89	34.48
	Gorra	96.97	55.68
	Bowl	6.19	12.58
	Taza 2	35.9	29.54
	Lata	0	0.01
	Caja	67.62	52.11

Profundidad	Obj	% seg.	% solap.
	Taza 1	92.16	66.29
	Gorra	92.67	56.85
	Bowl	63.83	40.60
	Taza 2	80.51	55.79
	Lata	10.63	12.34
	Caja	30.62	44.94

RGB-D	Obj	% seg.	% solap.
	Taza 1	90.05	65.07
	Gorra	90.82	42.83
	Bowl	68.54	42.88
	Taza 2	48.33	35.98
	Lata	23.67	18.81
	Caja	23.29	35.03

Conclusiones y trabajo a futuro

- conclusiones
- mejoras a implementar