

Detección de Motocicletas y del uso del casco de seguridad utilizando un algoritmo que emplea técnicas de procesamiento de imágenes y modelos de inteligencia artificial

ESP. ING. MARIO ANDRÉS VARÓN M.SC

MOVICI-MOYCOT 2018: JOINT CONFERENCE FOR URBAN MOBILITY IN THE SMARTY CITY

PAPER PRESENTATION

MARIO.VARON@PROTRAFICO.COM

Introducción

1. En Colombia los motociclistas son las primeras víctimas de la accidentalidad vial (Instituto Javeriano de Salud Pública, 2015).
2. Entre el 2007 y el 2014, se contabilizaron cerca de 28.000 motociclistas fallecidos. El aumento de la mortalidad de los motociclistas se debe:
 - I. Crecimiento desbordado del parque automotor, con un incremento del 445% en 14 años (El Espectador, 2015).
 - II. La falta de uso de elementos de protección pasiva, como el casco de seguridad.
3. No existe evidencia directa que permita inferir la proporción de la población de motociclistas que hace uso de sistemas de protección pasiva. Se han hecho estudios indirectos del análisis forense del occiso con posterioridad a la ocurrencia del accidente de tránsito. Se obtuvo que el uso es del 40% (Valbuena, 2012).

Problema

1. Para atender las problemáticas de seguridad vial, que se mencionan con anterioridad, se requiere la detección de forma precisa, ágil y a bajo costo de las motocicletas dentro del flujo vehicular, así como la detección del casco de seguridad por parte de aquellos que la conducen.
2. Se utilizan Sistemas Inteligentes de Transporte (ITS) para la detección de Motocicletas y el uso del casco de seguridad.
 - I. Sensores activos basados en tecnología LIDAR y DOPLER. La gran desventaja es que resultan ser muy costosos.
 - II. Sensores pasivos tales como Cámaras.
3. La detección de Motocicletas y del casco de seguridad, es bastante difícil. Ambos objetos presentan alta variabilidad en su forma, color y tamaño. Se añade a esto, ambientes de tránsito altamente congestionados, cambios en la iluminación y la interacción aleatoria entre los diferentes actores viales (se puede generar oclusión).

Estado del Arte

1. Detección, Reconocimiento y Clasificación de Motocicletas

- Más de diez estudios han abordado la detección de Motocicletas con cámaras. En la mayoría de estos casos, la cámara está estática. Se utiliza así mismo la sustracción de la imagen de fondo para la detección de los objetos en movimiento.
- Las distintas investigaciones hacen uso de elementos propios de las motocicletas, tales como elementos circulares, Histograma de Gradiente (HOG) para el color o la forma.
- La novedad en el estado del arte está en el uso de redes neuronales convolucionales. Un modelo muy poderoso de inteligencia artificial que es capaz de aprender de las características propias de las motos (esto requiere Datasets con tamaños de al menos un millón de imágenes).
- La precisión para la detección de motocicletas es del 97,8%.

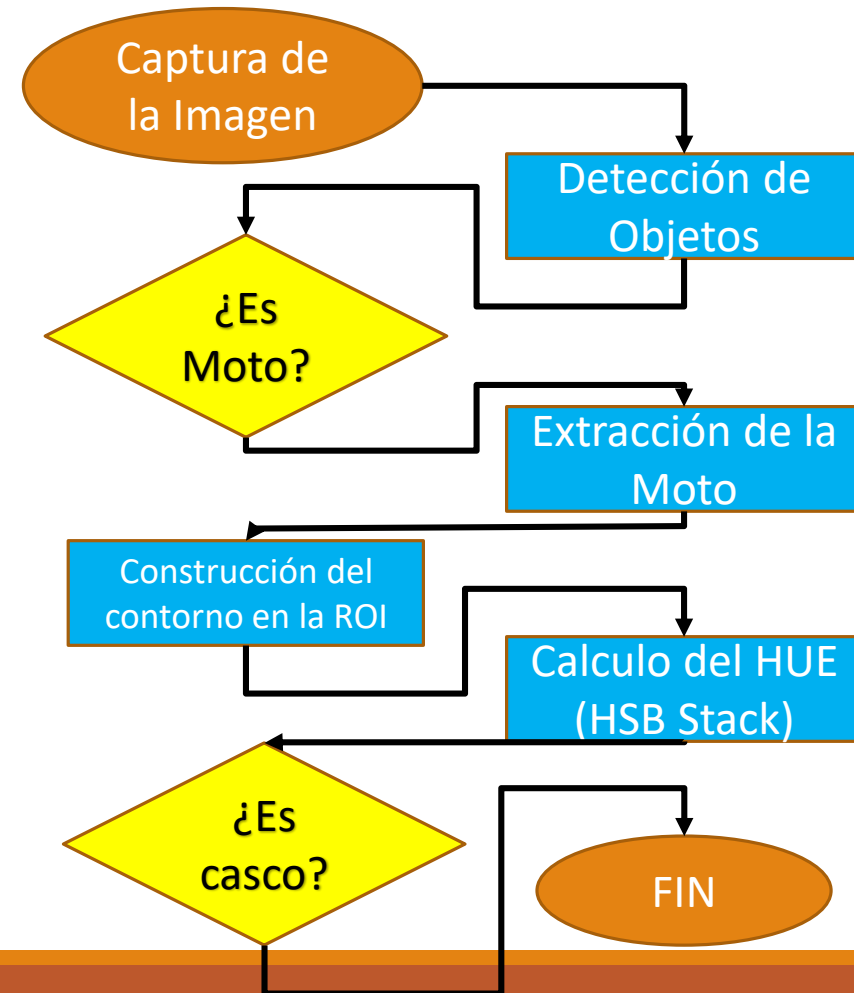
Estado del Arte

- Detección y Reconocimiento del uso del casco de seguridad en Motociclistas.
- Chiverton (2012): - Precisión del 85%.
 - ✓ Detección de un objeto en movimiento
 - ✓ Aislamiento de la región de interés. Cálculo de su histograma después de la aplicación de varios operadores.
 - ✓ Determinación del uso del casco por medio del análisis de características de la imagen transformada. Uso del brillo que genera el material del casco para determinar su presencia en la imagen. Estas características son utilizadas por un clasificador SVM (Support Vector Machine).
 - ✓ Seguimiento de la motocicleta haciendo uso de filtros de Kalman.



Metodología Propuesta

- Detección de objetos en el flujo vehicular.
- Clasificación binaria para identificación de las motocicletas.
- Ubicación y Extracción de la moto desde la imagen original.
- Extracción del ROI (región de interés) del motociclista y construcción del contorno de la zona de la cabeza del conductor, así como de la coordenada H en el stack HSB (hue, saturation and brightness) de la región de interés.
- Clasificación binaria de estas imágenes para identificación del uso del casco.



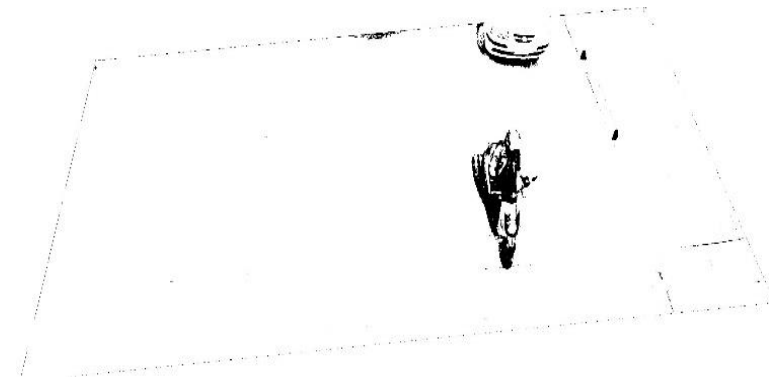
Geometría de la Captura

- Las imágenes se tomaron en un puente peatonal, a cuatro metros con diez centímetros (4,1 mts) de altura sobre la rasante de la vía, entre las 12 pm y 3 pm, en un día nublado, en la ciudad de Bogotá.
- Se utilizó una cámara GOPRO 4 de 12 MP sobre un trípode con una inclinación de -40° sobre el eje horizontal. El dataset original estuvo compuesto por 10.120 imágenes, de las que solo 820 tuvieron objetos en ella.
- La forma de captura propuesta disminuye las posibilidades de encontrar oclusión en los objetos que se presentan en ella, también supone un reto mayor, para la identificación apropiada de los objetos, en la medida en que puede existir una gran cantidad de ellos en un momento dado (i.e congestión vehicular).



Detección de Objetos

- Se sustrae la imagen del fondo.
- Se selecciona una ROI (Región de Interés dentro de la imagen).
- Se disminuye el contraste.
- Se convierte a escala de grises de 8-bits.
- Se umbraliza con la técnica de Momentos con el fondo invertido.



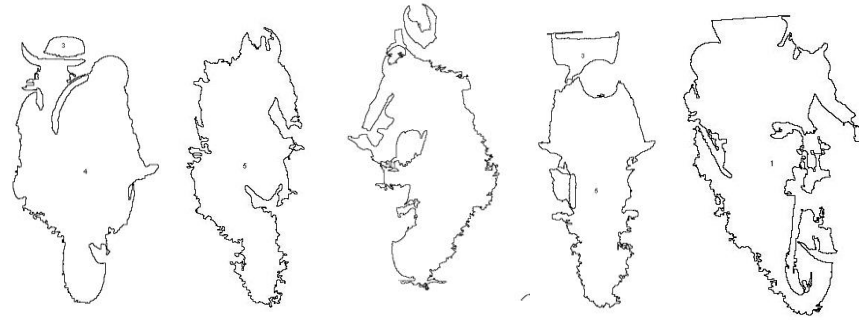
Detección de Objetos

- Se buscan regiones con morfología circular dentro de la imagen (análisis de partículas).
- Son importantes las regiones que ocupen entre el 5% al 10% del tamaño de la imagen (ese es el tamaño de la moto en la vía).
- El algoritmo construye la morfología del objeto utilizando el borde.
- Se extraen los objetos que se encuentran en la imagen.

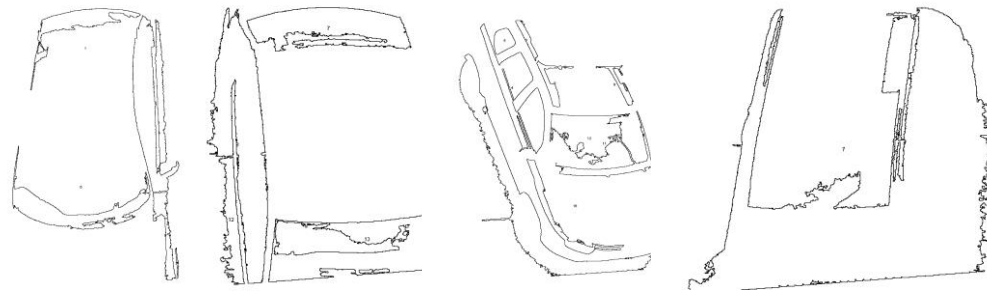


Ejemplos

Clase Moto



Clase No Moto



¿Es Moto?

- No se puede utilizar red neuronal convolucional propia (DS muy pequeño).
- Se utilizan tres modelos pre entrenados:
 - Iter_45
 - Inception-V3 Network
 - Full Image Network
- Se extraen un vector de las características propias de cada imagen según las capas de salida de cada modelo.
- El Dataset de Prueba es de 109 Imágenes. El de prueba es de 35.
- Se construyen varios clasificadores. La mejor precisión es con Boosted trees. Metaclasificador que utiliza “learners de base” y puede moldear relaciones no lineales. Precisión del 97,15%.

```
: {'accuracy': 0.9714285714285714, 'auc': 0.96875,  
   target_label    int  
   predicted_label int  
   count          int
```

Rows: 3

Data:

| target_label | predicted_label | count |
|--------------|-----------------|-------|
| 1 | 0 | 1 |
| 1 | 1 | 15 |
| 0 | 0 | 19 |

```
[3 rows x 3 columns], 'f1_score': 0.967741935483871
```

¿Es Moto?

- No se puede utilizar red neuronal convolucional propia (DS muy pequeño).
- Se utilizan tres modelos pre entrenados:
 - Iter_45
 - Inception-V3 Network
 - Full Image Network
- Se extraen un vector de las características propias de cada imagen según las capas de salida de cada modelo.
- El Dataset de Prueba es de 109 Imágenes. El de prueba es de 35.
- Se construyen varios clasificadores. La mejor precisión es con Boosted trees. Metaclasificador que utiliza “learners de base” y puede moldear relaciones no lineales. Precisión del 97,15%.

```
: {'accuracy': 0.9714285714285714, 'auc': 0.96875,  
   target_label    int  
   predicted_label int  
   count          int
```

Rows: 3

Data:

| target_label | predicted_label | count |
|--------------|-----------------|-------|
| 1 | 0 | 1 |
| 1 | 1 | 15 |
| 0 | 0 | 19 |

```
[3 rows x 3 columns], 'f1_score': 0.967741935483871
```

Extracción de Moto

- Se extrae la imagen de la moto utilizando un tamaño uniforme para todas las imágenes.
- Se suma nuevamente el fondo.
- Se construyen dos datasets:
 - Uno en RGB.
 - Otro con el fondo sustraído.
- Se selecciona el ROI de interés (tercio superior de las imágenes).



Construcción del contorno en el ROI

- Se descarta manualmente todas aquellas que tengan más de un conductor (oclusión de un casco sobre otro).
- Se convierte en escala de grises, la imagen con el fondo sustraído. Se umbraliza utilizando máxima entropía con fondo invertido.
- Se hace un llenado morfológico con su posterior cierre.
- Se aplica el operador de Sobel para la detección del borde.



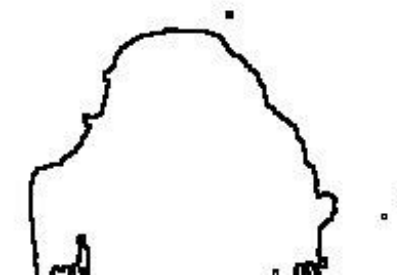
Construcción del contorno en el ROI Dataset Alternativo

- No hubo ningún motociclista sin casco en el Data Set.
- Se requirió la construcción de un DS Alternativo.
- Se construyó en la ciclovía en Bogotá con la misma geometría de captura que el de motos.
- Se aplica el mismo algoritmo.



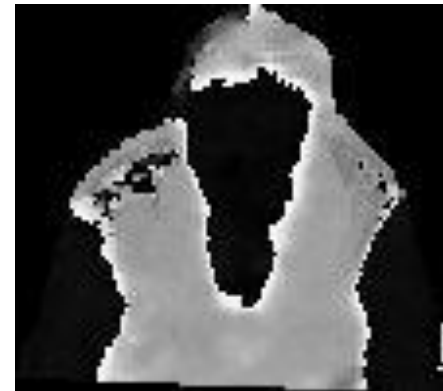
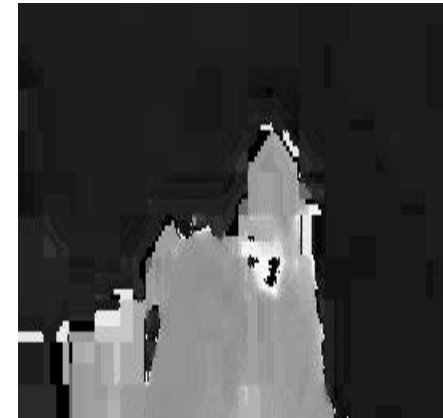
Cálculo de HUE (HSB Stack)

- La morfología no es un buen descriptor del uso del casco, ya que la geometría de la captura permite identificar motos en todas las direcciones y los ángulos. Por tanto resulta inservible.
- Se puede alimentar un clasificador con la foto en RGB, pero tomará mucho tiempo en converger. La precisión es baja por el tamaño del Dataset.
- No es posible utilizar momento o HOG, por la falta de elementos de comparación en el DS.



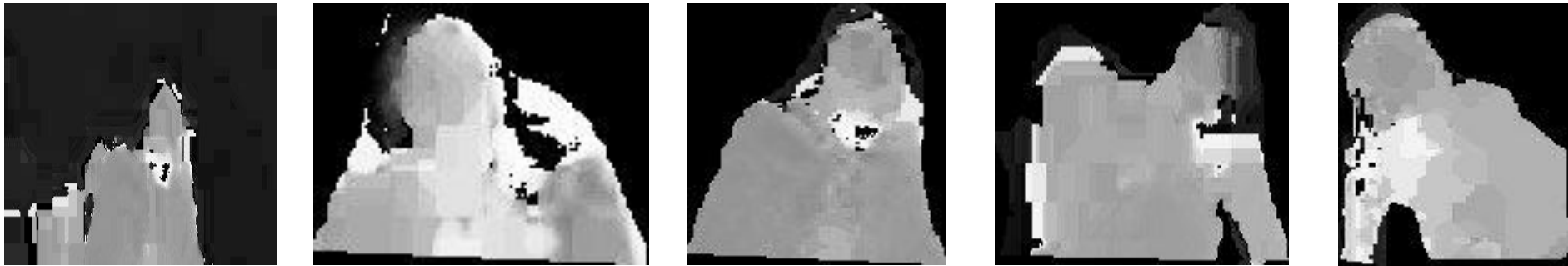
Cálculo de HUE (HSB Stack)

- La escala del color básico se representa con el valor en Hue del Stack HSB. Dicha escala es un valor angular. Sirve para determinar los cambios de color en las imágenes RGB.
- Se normaliza las distancia, tanto para el DS de bicicleta como para el DS de motocicleta. Solo se extraen objetos a distancia similar al punto de captura.
- Se utiliza el contorno para extraer la imagen del DS en RGB. Se vacía el contenido exterior.
- Los cambios de color en personas que utilizan el casco son menores.

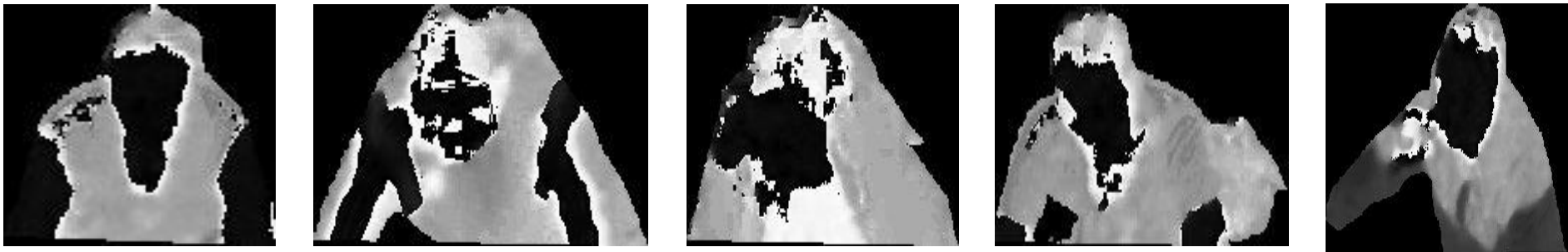


Ejemplos

Clase Con Casco



Clase Sin Casco



¿Es Casco?

- No se puede utilizar red neuronal convolucional propia (DS muy pequeño).
- Se utilizan los tres modelos pre entrenados:
 - Iter_45
 - Inception-V3 Network
 - Full Image Network
- Se extraen un vector de las características propias de cada imagen según las capas de salida de cada modelo.
- El Dataset de Prueba es de 77 Imágenes. El de prueba es de 37.
- Se construyen varios clasificadores. La mejor precisión es con Boosted trees. Metaclasificador que utiliza “learners de base” y puede moldear relaciones no lineales. Precisión del 85,29%.

```
{'accuracy': 0.8529411764705882, 'confusion_matrix':  
  target_label  int  
  predicted_label int  
  count  int
```

Rows: 4

Data:

| target_label | predicted_label | count |
|--------------|-----------------|-------|
| 1 | 1 | 18 |
| 1 | 0 | 2 |
| 0 | 1 | 3 |
| 0 | 0 | 11 |

```
[4 rows x 3 columns], 'f1_score': 0.8780487804878048,
```

Limitaciones

- El método no funciona si las personas portan cachuchas, gorros u otros elementos que no generen un cambio de color drástico con la cara. Para estos casos, el análisis por momentos invariantes de Hu o el uso de HOG puede ser útil. Se debe modificar la geometría de la captura y adoptar la propuesta por Chiverton (2014).
- Se debe normalizar la distancia. No funciona para distancias muy distintas frente el punto de captura.
- Se debe utilizar un Dataset de al menos 1000 imágenes para mejorar la precisión de ambos clasificadores.

Conclusiones

- La precisión para el estado del arte en detección de motos es del 97,8%. La metodología propuesta alcanza 97,14%.
- La precisión para el estado del arte en detección de casco de seguridad es del 89,13%. La metodología propuesta alcanza 85,29%.
- Investigaciones futuras deben intentar hacer uso de dos puntos de captura. Uno como el propuesto aquí y el otro como el que utilizó Chiverton. La conjunción de ambas arroja una precisión muy elevada.
- La metodología propuesta es novedosa, ya que la mayor parte del trabajo se hace en el procesamiento de imágenes y no en la red neuronal. Esto permite que los modelos extraigan las características de las imágenes rápidamente (<10 minutos en un portátil tradicional).
- Se debe resolver el problema de la identificación de pasajeros.

Gracias !

