

ARQUITECTURA PARA LA EXTRACCIÓN DE INFORMACIÓN DE ESPACIOS URBANOS

ARCHITECTURE FOR THE EXTRACTION OF INFORMATION OF URBAN SPACES

Mg. Alejandro Forero Guzmán
Pontificia Universidad Javeriana, Colombia
alejandro.forero@javeriana.edu.co

Ph.D. Carlos Parra Rodríguez
Pontificia Universidad Javeriana, Colombia
carlos.parra@javeriana.edu.co

(Artículo de **Investigación científica y tecnológica**. Recibido el 13/10/2011. Aprobado el 19/12/2011)

Resumen. El presente artículo presenta una arquitectura para la extracción de información de espacios urbanos, la extracción de información de espacios urbanos, la cual debe permitir la integración entre los sistemas de visión por computador que procesan imágenes o video con estrategias de bajo nivel, con la interpretación semántica que los humanos hacen de los espacios. En este documento también se presenta la estructura de esta arquitectura y las limitaciones que introducen los sistemas de adquisición con cámaras.

Palabras clave: Adquisición; representación de conocimiento; visión por computador.

Abstract. This paper presents an architecture for information extraction of urban spaces, which should provide integration between computer vision systems that process video images or low-level strategies, with the semantic interpretation that humans make of spaces urban. This document also presents the structure of this architecture and the constraints introduced by the camera acquisition systems.

Keywords: Knowledge acquisition/ representation, computer vision.

1. INTRODUCCIÓN

La cantidad de actividades que el ser humano realiza en las ciudades se incrementa y diversifica a medida que éstas crecen y se convierten en espacios más variados y, por ende, más complejos para describir y analizar. La movilidad es un ejemplo de acciones urbanas que se ha hecho más confusa y engorrosa con el crecimiento de las ciudades; el tráfico, los accidentes y la contaminación, son aspectos negativos que se acentúan con la masificación de la necesidad de desplazarse. La ciudad de Bogotá no ha sido ajena a esta situación, de hecho, el número de vehículos particulares registrados en la ciudad aumentó más del 60% entre 2008 y 2011 [1]; esta es una tendencia que se observa en todas las ciudades del mundo.

Afortunadamente, al tiempo que las necesidades surgen, también se generan alternativas tecnológicas que buscan mitigar dichas necesidades. Por ejemplo, los semáforos se implementaron para minimizar los problemas ocasionados por el tráfico. La incorporación de la tecnología pretende mantener bajo control las necesidades que crecen en las ciudades. Sin embargo, estas urgencias están interrelacionadas entre ellas: los problemas relacionados con la movilidad afectan la concepción misma de la ciudad, la forma en la que se planifica, se construye y se establecen regulaciones, no solo del tráfico sino de todos los aspectos de la vida urbana. Con esta perspectiva, las necesidades que se crean requieren más información. Y la cantidad de información debe crecer al menos al mismo ritmo con el que se efectúa cada actividad en la ciudad. Tener información suficiente permitirá planear y aplicar las medidas adecuadas para afrontar los requerimientos de la ciudad, entre ellos la movilidad o la seguridad.

Los requerimientos que plantea la movilidad, dio origen a la ubicación de unos sensores que tenían por objeto registrar la cantidad de vehículos que circulan por un lugar en un día. Los primeros sensores se desarrollaron con el principio de inducción electromagnética porque permitían registrar el paso de los vehículos; sin embargo, desde las últimas décadas, se ha privilegiado como sistema de adquisición de información las cámaras y procesamiento de video, en particular por la posibilidad de ampliar las variables para observar [2]. En la actualidad, tan importante como la información de los vehículos que circulan, es la información de los peatones y demás usuarios de la vía. Esto es especialmente notorio en espacios distintos a las autopistas, en las que la diversidad de

usos y usuarios de las vías hace que los sistemas basados en visión sean ideales para extraer la información de lo que sucede en estos espacios urbanos [3]. En la Fig. 1 se ilustra un ejemplo de la versatilidad de los sistemas sustentados en visión, (imagen tomada de [4]), donde simultáneamente se extrae información de los vehículos que circulan, se identifican, se siguen y se estima la velocidad; al mismo tiempo, se detectan y siguen los peatones y grupos de peatones, para luego medir distancias entre peatones y vehículos.

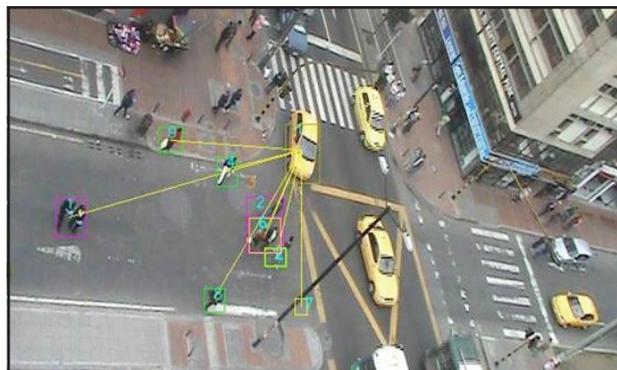


Fig. 1. Detección y seguimiento de múltiples variables por un sistema de Visión por computador

Hoy, los sistemas de cámaras y procesamiento de video se han convertido en sistemas de Visión por computador, encargados de extraer información del entorno, en este caso de los espacios urbanos [5]. La Visión por computador es el sensor privilegiado en la observación de las condiciones en la ciudad, el medio de adquisición de información para sistemas que tienen una complejidad mayor y que pueden incluir los sistemas de control de tráfico, de monitoreo urbano o de información de las condiciones existentes.

El avance de los sistemas de Visión por cámara, como sistema de adquisición de información de las condiciones observadas, se ha favorecido por los desarrollos en las capacidades de procesamiento en los sistemas de cómputo y por el creciente interés en los sistemas de visión en tiempo real. Sin embargo, no todos los avances realizados en estos campos tienen presente la naturalidad de la información que espera un usuario humano: “...there is still a big gap in analytical skills between a typical security guard and state-of-the-art image-processing algorithms” [6].

Esta tendencia de interpretar la información desde el procesamiento de imágenes sin integrar la interpretación de la información del lenguaje humano, produ-

ce dos problemas. Primero, aumenta la complejidad algorítmica para interpretar la imagen por no usar la información implícita que se tiene de la escena analizada. Por ejemplo, en la interpretación de una vía, varios de los sistemas de Visión por cámara no incorporan la percepción del conductor o del peatón de esa vía: como podrían ser la presencia de usuarios u obstáculos habituales.

Segundo, crea una ruptura entre los resultados generados por la interpretación de los sistemas de Visión por cámara, que se enfocan en el procesamiento de la imagen, y los usos y usuarios humanos, quienes precisan de esa información; esto es lo que algunos autores llaman “*semantic gap*” [7]. Visto de otra forma, estos despliegues no integran un aprendizaje de máquina que permita caracterizar los elementos del ambiente humano [8].

Dado que desde múltiples disciplinas se estudia, se planea o se ejerce algún control del espacio urbano y la forma de entender este espacio es en sí muy diversa, este trabajo presenta una arquitectura para adquirir información del espacio urbano a partir de la información entregada automáticamente por los sistemas de Visión por cámara. Esta estructura permitirá describir los fragmentos del espacio urbano a través de la interpretación de bajo nivel que hacen los sistemas de Visión por cámara, analizar píxeles o conjuntos de éstos para construir una descripción de la escena. Adicionalmente, la estructura incluye, en la descripción del espacio urbano, la información semántica que describe el espacio y el uso que se le da. Trabajos similares proponen una estructura topográfica y otra topológica, como [9] [10]. Esta idea se incorpora en la arquitectura que se propone.

Mediante la estructura con la que se describe el espacio urbano desde el sistema de Visión por cámara, se sugiere una arquitectura para almacenar la información de una escena, que se asocia con el espacio observado por una cámara de video. Además, en este trabajo, se definen las reglas encaminadas a conectar fragmentos del espacio urbano observado y cada fragmento sintetizado con la arquitectura propuesta. A largo plazo, su construcción debería permitir llegar a un modelo de un espacio urbano de tamaño arbitrario. Para tal propósito se parte de la descripción de múltiples escenas y de la síntesis de la información en la arquitectura que se propone.

Adicionalmente, si se enriquece la captura de información con las actividades en los espacios urbanos,

se plantea realimentar la estructura de datos propuesta con la descripción semántica de actividades del espacio observado. Así se incorporará a la arquitectura la información semántica que caracteriza el uso de un espacio urbano; un ejemplo de esta información es la descripción que se hace de una vía vehicular: el número de carriles que la componen, el tipo de tráfico que la recorre, el sentido del tráfico (p.ej. sur-norte), si la cruzan otras vías, o la velocidad máxima permitida, entre otros datos que se pueden asociar a la descripción de este espacio.

En la siguiente sección se formaliza el modelo con el que se describen las escenas. Después se detalla la arquitectura propuesta para almacenar la información de la escena desde la información de bajo nivel, luego, se describe la información semántica y, por último, se hace la integración entre las dos. En la cuarta sección se alude a los alcances y limitaciones de la arquitectura. Finalmente, se presentan las conclusiones de este trabajo.

2. FORMALIZACIÓN DEL MODELO DE LA ESCENA

El espacio urbano se considera como una unión de infinitud de espacios. De igual forma, todo espacio urbano está conformado por la unión de distintos elementos básicos. La descripción de un espacio urbano se construye desde la definición de los elementos que lo componen.

A partir de la definición de los elementos básicos que componen el espacio urbano se construye una representación de cualquier espacio que es observado en una ciudad. Esta idea es particularmente útil en la generación de un modelo que permita describir la ciudad a través de un sistema de Visión por computador (VC).

Al considerar el espacio urbano capturado en el plano imagen por una cámara de vista superior, se busca el conjunto de elementos básicos que se encontrará en el plano imagen, independiente del espacio urbano que se observa.

Dado que se intenta describir la dinámica que se observa en los espacios urbanos, se busca que los elementos básicos usados también sean relevantes para detallar las acciones que suceden en dicho espacio.

Teniendo presentes los criterios mencionados se tiene:

Sea E el conjunto de las distintas escenas que se capturan con el uso de una cámara de vista superior fija en el espacio, de forma que E_n , es una escena capturada y es un elemento de E .

El plano imagen que representa esta escena E_n , está conformado por m segmentos.

El conjunto de segmentos S , está compuesto por instancias de los elementos: vía (v), intersección (i), plaza (p) u obstáculo (o).

$$E_n \in \mathbf{E}$$

$$E_n = \bigcup_{j=1}^m (S_j)$$

$$S_j \in \mathbf{S}$$

$$\mathbf{S} = \{v, i, p, o\}$$

Los segmentos vía (v), se conforman como espacios de tránsito y se clasifican de acuerdo con los móviles que los recorren. Adicionalmente, se caracteriza por su estructura geométrica y las del tráfico que circula por el segmento, es decir, la velocidad y la densidad de los móviles.

Los segmentos intersección (i), son espacios en los que se encuentran dos o más vías y pueden clasificarse de acuerdo con los móviles que se encuentran en la intersección o con la estructura geométrica de la misma intersección.

Los segmentos plaza (p), se conforman como espacios de tránsito y/o estadía, primordialmente de peatones, que no se encuentran atravesados por una vía.

Los segmentos obstáculo (o), incluye cualquier segmento que no se clasifique dentro de uno de los tres segmentos anteriores. Estos segmentos son indispensables para agrupar a todos los segmentos sobre los que no es válido ningún desplazamiento de los móviles en la escena.

3. ARQUITECTURA PROPUESTA

Esta arquitectura sintetiza la interpretación del espacio urbano con el modelo que se plantea e introduce otros conceptos a partir de la interpretación de bajo nivel realizada por un sistema de Visión de cámara. Adicionalmente, intenta una estructura para enriquecer la descripción con la interpretación semántica del

espacio urbano observado basado en el uso que se hace de dicho espacio.

En general, cualquier sistema de Visión de cámara suele concentrarse en alguna característica semántica de una escena y trata de optimizar la identificación o extracción de la misma. La arquitectura que se propone no difiere de este principio básico, pero busca generalizarlo e incluir las actividades que se desarrollan en el espacio urbano o, al menos, producir una estructura capaz de incluir las actividades que se observan: datos básicos como el conteo de móviles que se desplazan (p.ej. vehículos o peatones) más elaborados como el registro de infracciones en un cruce semaforizado.

Con esta arquitectura se describe cómo está construido un espacio urbano, pero, además, se incluye la información de las acciones que se dan en ese espacio como la circulación de vehículos y los aspectos que caracterizan esa circulación, como las velocidades permitidas, el sentido de la circulación y demás normas que regulen esa actividad. A continuación se presenta cómo, desde la interpretación del plano imagen por un sistema Visión de cámara y la interpretación semántica del espacio se construye la arquitectura para extraer y almacenar información del espacio observado.

3.1 Interpretación de bajo nivel

La interpretación de bajo nivel está inspirada en los trabajos previos realizados en el grupo de investigación SIRP (Sistemas Inteligentes Percepción y Robótica) de la Pontificia Universidad Javeriana - Bogotá, en los que se han desarrollado técnicas de Visión de cámara relacionadas con la movilidad. Estos trabajos abarcan desde la detección del camino para una plataforma móvil [11], pasando por la adquisición de variables de tráfico en una escena particular [12], [13], [14], [15], [16], hasta la interacción de variables en la escena [16], [17].

En la interpretación del Plano imagen se genera un modelo del fondo de la escena observada [12], [18], [19]. Una vez estimado el modelo del fondo, se identifican los objetos que lo componen y se usa la información que suministra el usuario humano. Esta información corresponde a los objetos que físicamente conforman el espacio urbano observado. El modelo del fondo también se emplea para extraer el primer plano, el cual se usa para identificar los móviles que

recorren la escena. Para caracterizar los conjuntos que conforman la escena y los móviles se utilizan características basadas en HOG [20] y matrices de covarianza [21] que, aunque probados para la detección de peatones, por sus características, permiten identificar otros tipos de estructura [22], en especial cuando la información de los bordes es relevante.

Con el modelo de fondo se construye un nuevo Plano imagen, que se segmenta para distinguir los objetos que hacen parte del fondo. La técnica seleccionada para la segmentación debe permitir separar las vías, los edificios, la vegetación y una cuarta clase que agrupe a los demás objetos que se presenten en la escena, como el equipamiento urbano, entre otros.

El primer plano se segmenta para individualizar los móviles que transitan la escena. Posteriormente, los móviles deben ser clasificados, para la que se inicia con una tipificación dada por el tamaño de los móviles [2], [15], que se depura a medida que las capacidades de los algoritmos implementados y la calidad de los datos se vean mejorados. En el caso de peatones que caminan juntos, se clasifican como grupo de peatones [4].

En la Fig. 2 se presenta la estructura que conforma el núcleo de la arquitectura. Con el Plano imagen se construyen dos representaciones, el Fondo y el Primer plano. A su vez, cada una de estas representaciones se desglosa para agrupar los componentes que se presentan en la escena capturada en el Plano imagen.

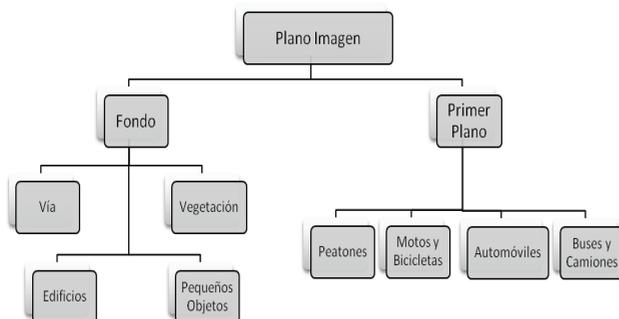


Fig. 2. Representaciones a partir del plano imagen

La representación del Fondo, se descompone en las clases que suelen componer la escena urbana, y de esta forma se construye una Clase vía y una Clase edificios (primera bifurcación en la Figura 2), que se encuentran en la mayoría de las escenas urbanas

que se observan desde la vista superior en ciudades como Bogotá. También, se adiciona una representación de la vegetación, como es observada por la cámara y extraída por el sistema de Visión de cámara, con lo que se construye la Clase vegetación. Se añade una representación que agrupa los objetos que completan la escena, como el equipamiento urbano. Esta representación puede incluir también pequeños objetos no previstos inicialmente y que hacen parte del fondo de la escena; con todo esto se construye la cuarta clase debajo de la Clase fondo.

De igual forma, se procede con la Clase Primer plano para descomponerla en cuatro clases, con el uso del tamaño de los objetos segmentados como primer criterio para agrupar los móviles detectados, tal como se muestra en la rama derecha de la Fig. 2. Otros criterios se pueden construir con la observación en el tiempo de los móviles.

3.2 Interpretación semántica

La interpretación semántica que el usuario humano construye del espacio urbano cuando se crean modelos cognitivos de apropiación de la información [23], se utilizará para enriquecer la descripción de la escena observada por la cámara [8], [24]. La arquitectura propone representar en forma de árbol la información semántica para facilitar el mapeo con la información extraída por el sistema de Visión de cámara. El árbol como estructura también permite incorporar la clasificación en conjuntos disyuntos que se hace tanto por la interpretación semántica como por el sistema de Visión de cámara.



Fig. 3. Bifurcación de la interpretación semántica

Se generan dos clases: la Clase escena, correspondiente al espacio urbano observado, y la Clase móviles, que corresponde a los móviles que recorren dicho espacio. Así, la Clase escena se corresponde con el Fondo extraído por el sistema de Visión de cámara; al igual que la Clase móviles que se corresponde con el Primer plano. Esta primera bifurcación se ilustra en la

Fig. 3, desde la que se intuye la similitud con la interpretación de bajo nivel propuesta en la sección anterior.

Para aprovechar las similitudes entre la interpretación semántica y los resultados del sistema de Visión de cámara, tanto la Clase escena como la Clase móviles, se subdividen en cuatro clases. Con esto se busca favorecer la correspondencia entre la interpretación del Plano imagen representado por el Fondo y el Primer plano, con la descripción semántica que se hace del espacio y los móviles que la recorren.

Por ejemplo, la Clase vía especifica un subconjunto de los objetos que conforman el fondo de la escena en la interpretación de bajo nivel. La información semántica que enriquece la descripción de esta clase incluye: el número de carriles que tiene la vía, el tipo de tráfico que la recorre, el sentido del tráfico (p.ej. sur-norte), si la cruzan otras vías, la velocidad máxima permitida, entre otros datos que definen el espacio o los usos que se le dan. La Fig. 4 ilustra este ejemplo.

Adicionalmente, a los detalles particulares de la clase que se construye, las características generales de la escena y la ubicación de la cámara complementan la información que se busca agrupar para completar la descripción de la escena. Esta descripción se puede continuar en más subniveles, siempre que el sistema de Visión de cámara sea capaz de diferenciar las subclases que se incluyan.

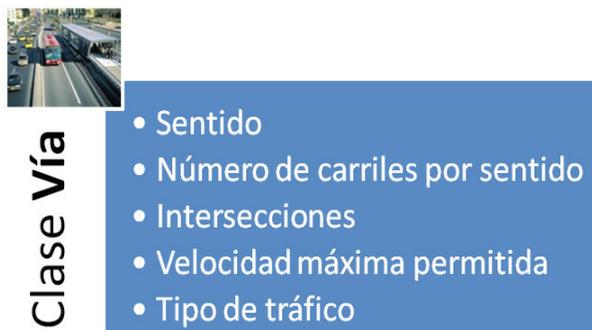


Fig. 4. Información semántica que enriquece la Clase vía

3.3 Integración de bajo nivel con la información semántica

La arquitectura que se propone busca enriquecer los datos que se extraen del plano imagen mediante el sistema de Visión de cámara, con la informa-

ción semántica que describe la escena. El objetivo es que cada una de las clases que conforman la interpretación de bajo nivel se enriquezca con la información semántica que se asocia con cada clase. Adicionalmente, la integración de estos datos se sintetiza en un repositorio que almacena la información que se extrae a medida que el sistema de adquisición de información funciona. De esta forma, se construye el conocimiento explícito que describe la escena, desde el nivel semántico hasta el nivel de procesamiento en el Plano imagen. En la Fig. 5 se observa un ejemplo para la Clase vía de cómo se integran los subsistemas y se almacena la información: en el eje vertical se depura un proceso de bajo nivel, mientras el eje horizontal adiciona la información semántica e integra el Repositorio de datos, donde se almacenan todos los descriptores de la clase.

El proceso descrito con la Clase vía se repite para cada una de las clases que componen la arquitectura, hasta alcanzar el nivel de detalle buscado en la descripción de la escena; el mismo procedimiento se sigue para completar la descripción de los móviles. De esta forma, se integra un esquema de aprendizaje que permite aprender los significados implícitos en el espacio para robustecer el sistema, al tiempo que se construye un lenguaje común con el usuario humano.

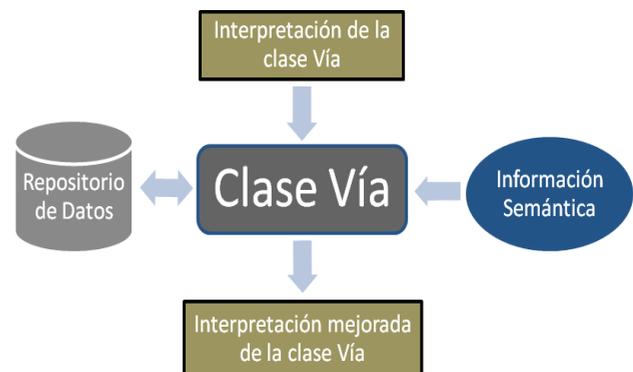


Fig. 5. Integración del procesamiento de bajo nivel con la información semántica y el Repositorio de datos.

Después de completar una estructura capaz de almacenar la información que describe la escena, se integra la información temporal que se produce con el desplazamiento de los móviles. La información temporal surge del procesamiento de los *frames* que conforman el video de observación. Así, los datos provenientes del sistema de Visión de cámara son producto de las posiciones de los móviles en cada uno de los planos imagen procesados (uno por cada

frame) y una etapa de seguimiento que reconstruye la trayectoria de los móviles.

La información temporal también se agrupa en niveles de abstracción que permite integrar la información semántica a los datos que se generan en la etapa de seguimiento del sistema de Visión de cámara. Así, se integra la información semántica que hace referencia a las posibles interacciones entre los móviles y la escena [24]. En el Plano imagen esta información ayuda a especificar las oclusiones que se pueden dar por la posición de la cámara, la estructura de la escena y la ruta de desplazamiento de los móviles.

Esta información temporal se agrupa en la Clase interacciones que conecta la información condensada en la Clase escena y en la Clase móviles, y permite un nivel de abstracción de la información extraída más allá de las posibilidades del sistema de Visión de cámara, puesto que da elementos para interpretar los datos capturados, aun en la eventualidad de oclusiones o ante posibles errores del sistema.

Una vez definida la escena, registrados los móviles y adicionada la información de las interacciones posibles entre móviles y móviles, y entre móviles y escena [2], se construye un registro del desplazamiento de cada móvil en la escena. Esta información adicional se condensa en una definición de eventos [2]. La Clase eventos es el último nivel de información que se almacena en la arquitectura que extrae la información de la escena. Esta clase condensa la información espacial y temporal de bajo nivel y la transforma en argumentos semánticos para ser interpretados por un usuario humano, tales como contravenciones de tránsito [25], accidentes, situaciones de riesgo, problemas de seguridad y, en general, descripciones del comportamiento de las personas que habitan los espacios urbanos y sus objetos [2], [6]. En la Fig. 6 se presenta una gráfica que ilustra cómo interrelacionan las Clases Escena, Móviles e Interacciones.

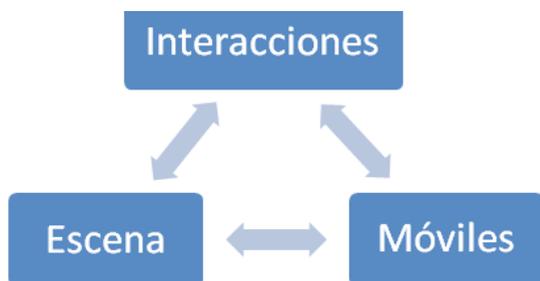


Fig. 6. Relación entre las clases Escena, Móviles e Interacciones

Las interacciones se constituyen en un tercer eje que agrupa la información en distintos niveles de abstracción hasta definir la Clase eventos, tal como se observa en la Fig. 7.

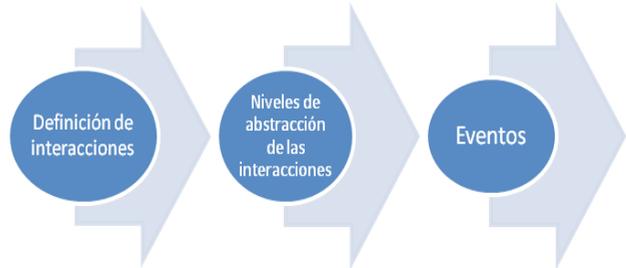


Fig. 7. De la definición de interacciones a la construcción de eventos

4. LIMITACIONES Y ALCANCES DE LA ARQUITECTURA

Las capacidades del sistema de adquisición de información de escenas urbanas, basado en la arquitectura propuesta, están limitadas por la observabilidad del espacio y la observabilidad de los eventos que se registren sobre cada espacio.

En cuanto a la observabilidad del espacio, la escena que captura el sistema de Visión de cámara y en particular el Plano imagen que se genera, está limitado por el tamaño mismo de la escena o por su región visible; por las características del sensor y de la óptica de la cámara; y por la ubicación misma de la cámara, en relación con el espacio que se observa. Dada la infinidad de combinaciones que se producen entre estos tres elementos, (escena, características de la cámara y ubicación relativa entre la cámara y la escena), las limitaciones del sistema se describirán sobre las características del Plano imagen, que es el resultado final de toda la etapa de captura.

Para el correcto funcionamiento del sistema de Visión de cámara se debe trabajar con un mínimo de información cuantificada por la cantidad de píxeles por objeto observado. También se deben tener en cuenta las oclusiones que se dan entre los objetos observados, pues éstas también limitan la cantidad de píxeles asociados con cada objeto. En últimas, se pretende definir un mínimo para que el sistema de Visión de cámara distinga los objetos en el Plano imagen y posteriormente sea capaz de seguir dichos objetos en el tiempo mientras recorren la escena.

Siguiendo esta idea, se propone trabajar con imágenes de 640 por 480 píxeles y limitar el espacio observado a 10m x 7,5m. Con esto se tendría un mínimo de 64 píxeles por metro lineal. (64 píxeles x 64 píxeles por m²). Con esta limitación del espacio, se podrían tener espacios más pequeños porque se aumenta la resolución. Esta limitación está dada para observar peatones, puesto que, en el caso de que el espacio sea únicamente vehicular, la resolución en píxeles podría bajar hasta 4 veces, es decir, observar espacios de hasta 40m x 30m y mantener la misma resolución en píxeles.

Si este espacio se toma sobre una vía vehicular, la cantidad de carros se limitaría a los que se observan, de acuerdo con el número de carriles de vía. También se propone mantener la estructura 4 a 3 del encuadre del plano observado. En cuanto a la densidad vehicular en la vía, la cantidad de vehículos presentes estará limitada por el tamaño de los mismos, teniendo presente que el tamaño promedio de un automóvil es 1,8m x 4,5m.

Si se observa un espacio peatonal, la evaluación del modelo semántico se limitará a espacios no atiborrados de transeúntes. En términos cuantitativos, sobre el espacio de 70 m² se espera que no haya más de 35 transeúntes simultáneamente, ($A [m^2]/2$) y que en ninguna parte se encuentren más de 3 personas por metro cuadrado. Estas limitaciones se plantean para concentrarse en la descripción semántica y no generar una dependencia del desempeño de los algoritmos de Visión de cámara de bajo nivel.

5. CONCLUSIÓN

Con este trabajo se busca avanzar en el despliegue de sistemas de adquisición de información en forma automática, con los que se enfrentan dos de los problemas que más preocupan a la sociedad en las grandes ciudades, como son: la seguridad y la movilidad. Sin embargo, la arquitectura propuesta plantea un sistema de adquisición de información que incluye las ventajas de los sistemas automáticos pero sin dejar de lado la interpretación humana de esa información. Desde esta perspectiva, la arquitectura debería servir para tomar decisiones sobre el uso de los espacios urbanos.

REFERENCIAS

- [1] Observatorio Ambiental de Bogotá. Secretaría Distrital de Ambiente. "Número de vehículos particulares-NVP" [Online]. Available: <http://oab.ambientebogota.gov.co/index.shtml?s=l&id=272&v=l>. Nov. 10, 2011.
- [2] P. Kumar, S. Ranganath, H. Weimin, and K. Sengupta, "Framework for Real-Time Behavior Interpretation From Traffic Video," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, no. 1, pp. 43-53, Mar. 2005.
- [3] C. Machy, C. Carincotte, and X. Desurmont, "On the use of Video Content Analysis in ITS : A review from academic to commercial applications," in *2009 9th International Conference on Intelligent Transport Systems Telecommunications, (ITST)*, 2009, no. 23, pp. 574-579.
- [4] A. Forero, C. Diaz, A. Sanchez, L. Quintana, F. Mariño, and L. Barrero, "Implementación de una Herramienta para la Adquisición de Variables de Tráfico: Peatones y Vehículos, Utilizada en la Evaluación del Riesgo de Peatones que Cruzan la Vía," in *XIII STSIVA*, 2008, pp. 95-98.
- [5] C. Regazzoni, A. Cavallaro, Y. Wu, J. Konrad, and A. Hampapur, "Video Analytics for Surveillance: Theory and Practice [From the Guest Editors]," *IEEE Signal Processing Magazine*, vol. 27, no. 5, pp. 16-17, Sep-2010.
- [6] J. Candamo, M. Shreve, D. B. Goldgof, D. B. Sapper, and R. Kasturi, "Understanding Transit Scenes: A Survey on Human Behavior-Recognition Algorithms," *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 1, pp. 206-224, Mar. 2010.
- [7] R. Porter, A. Fraser, and D. Hush, "Wide-Area Motion Imagery," *IEEE Signal Processing Magazine*, vol. 27, no. 5, pp. 56-65, Sep-2010.
- [8] D. Crevier and R. Lepage, "Knowledge-Based Image Understanding Systems : A Survey," *Computer Vision and Image Understanding*, vol. 67, no. 2, pp. 161-185, 1997.
- [9] B. T. Morris and M. M. Trivedi, "Trajectory learning for activity understanding: unsupervised, multilevel, and long-term adaptive approach.," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 11, pp. 2287-301, Nov. 2011.
- [10] D. Makris and T. Ellis, "Learning semantic scene models from observing activity in visual surveillance.," *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics : a publication of the IEEE Systems, Man, and*

- Cybernetics Society, vol. 35, no. 3. pp. 397–408, Jun-2005.
- [11] A. Forero and C. Parra, "Extraction of Roads From Out Door Images," in *Vision Systems: Applications*, no. June 2007, G. Obinata and A. Dutta, Eds. InTech, 2007, pp. 101-112.
- [12] G. E. Urrego, F. C. Calderón, A. Forero, and J. A. Quiroga, "Adquisición de variables de Tráfico vehicular usando visión por computador," *Revista de ingeniería Universidad de los Andes*, no. 30, pp. 7-15, 2009.
- [13] J. Quiroga, N. Romero, C. García, and C. Parra, "Adquisición de variables de tráfico peatonal utilizando visión por computador Pedestrian traffic variables acquisition using computer vision," *Revista Facultad de Ingeniería Universidad de Antioquia*, no. 60, pp. 51-61, 2011.
- [14] D. Insuasti, J. Quiroga, and A. Forero, "Detección y Seguimiento de Vehículos Automotores en Video," in *XIII STSIVA*, 2008.
- [15] F. C. Calderón and A. Forero, "Low Complexity Algorithm For The Extraction Of Vehicular Traffic Variables," in *IEEE ITSC 2011*, 2011.
- [16] E. A. Romero, J. Quiroga, and A. Forero, "Seguimiento de Personas Basado en los Descriptores HOG," in *XVI STSIVA*, 2011, no. 2.
- [17] A. Forero, J. Quiroga, C. Parra, and C. L. Niño, "Información y caracterización del tráfico vehicular usando visión por computador," in *Reunión nacional ACOFI 2011*, 2011.
- [18] M. David and J. Quiroga, "Estudio y Análisis del Algoritmo de Sustracción de Fondo Codebook," in *XV STSIVA*, 2011.
- [19] K. Kim, T. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground-background segmentation using codebook model," *Real-Time Imaging*, vol. 11, no. 3, pp. 172-185, Jun. 2005.
- [20] N. Dalal and B. Triggs, *Histograms of Oriented Gradients for Human Detection*. IEEE, 2005, pp. 886-893.
- [21] F. Porikli, P. Meer, and O. Tuzel, "Human detection via classification on riemannian manifolds," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2007, pp. 1-8.
- [22] F. M. Porikli, "Classification on Riemannian Manifolds," *videlectures.net*, 2010.
- [23] J. M. Zacks and B. Tversky, "Event structure in perception and conception.," *Psychological bulletin*, vol. 127, no. 1, pp. 3-21, Jan. 2001.
- [24] C. Hudelot, "Towards a Cognitive Vision Platform for Semantic Image Interpretation," *UNIVERSITE DE NICE - SOPHIA ANTIPOLIS*, 2005.
- [25] J. A. Vijverberg, N. A. H. de Koning, J. Han, P. H. N. de With, and D. Cornelissen, "High-Level Traffic-Violation Detection for Embedded Traffic Analysis," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07, 2007*, Vol. 22, No. 12, p. 11-793-11-796.