

Hacia una Instantánea Urbana. Captura, Lectura y Manejo Automático de Información en Forma de Personas a Partir de Registros Videográficos a Nivel de Observador.

Towards an Urban Instant. Automatic Reading of Information in the Shape of People at Observer Level.

Claudio Araneda Gutiérrez.

Universidad del Bio-Bío.

claraneda73@gmail.com

Braulio Gaticas Laurie.

Universidad del Bio-Bío.

bgaticalaurie@gmail.com

Abstract

People detection remains a challenging field within computer vision, the focus being placed mainly on the quantification of discreet amount of people. This paper reports on the progress related to the Urban Instant Project and deals specifically with a first attempt at automatically detecting and quantifying the continuous amount of people present in any given video record of the same format taken from a walking observer point of view. In order to do so, we use the optical flow approach to image reading. This yields a first quantification of information in the shape of people within the total field of vision.

Keywords: Urban Instant; People Detection; Computer Vision; Observer; Phenomenology.

Introducción

La detección automática de personas a partir de registros de video constituye un campo emergente y en tanto tal, aun no exento de dificultades y desafíos pendientes. Importante entre ellos, aquel que dice relación con el problema de la oclusión de personas por otras personas, lo que acarrea dificultades a la hora de cuantificar la cantidad discreta (número) de personas presentes en un registro dado (Tang et al. 2012, 2013, 2014; Dollár et al. 2010, 2012). Así, los registros más exactos obtenidos en este campo han sido aquellos efectuados por cámaras situadas en altura tales como cámaras de seguridad y vigilancia (Ratti, 2012). A diferencia y en complemento a esta prolífica línea de investigación, el proyecto Instantánea Urbana tiene como objetivo central desarrollar una estrategia de detección, lectura y cuantificación automática de la cantidad continua de personas presentes en el campo visual fotográficamente abstraído, desde el punto de vista del peatón promedio. Es decir, a la altura de un observador caminando.

En este contexto analítico, la oclusión no es un problema a resolver sino una variable integral a la metodología planteada. En efecto, el foco lo constituye el comportamiento de lo que denominaremos "información visual en forma de personas" para diferenciarla de toda otra información visual contenida en el total del campo visual promedio de un peatón tal como podría ser información visual "en forma de naturaleza" o bien información visual "en forma de espacio urbano". Cada uno de estos tipos de información constituye un porcentaje del total del campo visual fotográficamente abstraído y en conjunto conforman un 100% que hemos denominamos "campo visual arquetípico" (Araneda 2013^a, 2013b, 2014). Con el objetivo de aislar y cuantificar la

variable "personas" del total del campo visual digitalmente abstraído recurrimos a la estrategia del "flujo óptico" la que nos permite detectar de forma automática todos aquellos cuerpos o elementos en movimiento en una imagen de video. El experimento arroja luz por vez primera sobre el comportamiento de esta "nube de gente" en permanente metamorfosis así como de nuestro campo visual en relación a la presencia y ausencia de personas. Este trabajo se inserta en el contexto del proyecto Instantánea o Polaroid Urbana (Araneda 2013b, 2014) e introduce una primera aproximación al levantamiento, lectura y manejo sistemático y automático de información visual en forma de personas.

Aspectos Metodológicos.

Se trabaja con tres segmentos de video de 60 segundos tomados en tres puntos de la calle peatonal mas transitada de la ciudad de Concepción. El formato del registro queda determinado por las siguientes consideraciones. Cámara fija (Smartphone Sony Xperia) a altura de observador dispuesta en el centro del espacio peatonal sobre monópode. La resolución del registro fue de 1280 x 960 y fue procesado a 680 x 480 pixeles. La captura fue realizada un día de semana a las 13.00 horas.

El registro fue procesado en plataforma rhinoceros a través del plug-in grasshopper mediante una programación dedicada. El procesamiento de visión artificial se realiza con ayuda del paquete de software "firefly". Para determinar un mapa vectorial se ocupa específicamente el algoritmo "optical flow", el que permite determinar de forma automática los porcentajes de área de imagen correspondiente a personas por *frame*, proceso que en los dos pilotos anteriores fue

realizado “rasterizando” de forma análoga toda superficie que formase parte de personas.

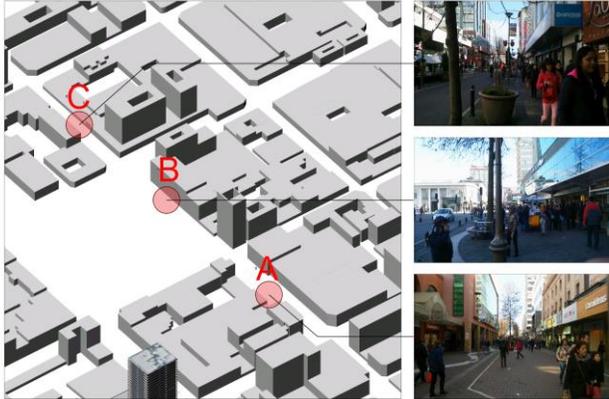


Figura 1: Posición de las muestras dentro de la trama urbana,

Estos registros son posteriormente estandarizados bajo un formato universal consistente en geo-referenciar los puntos muestreados y la determinación del campo visual del punto de observación, dejándolo en un lenguaje matemático utilizable. Se abre así la posibilidad de compatibilizar este trabajo con otras plataformas GIS estándares tales como api de Google Street View. Esto abre a su vez la posibilidad de realizar estudios fotométricos, contrastando cada *frame* del registro con una modelación 3D de la ciudad integrada a nuestra programación para determinar la posición, orientación y apertura del campo visual capturado por la muestra.



Figura 2: Fotogramas de las muestras A-B-C cada 10 seg

Con esto, se procede a registrar la presencia de gente en el total del registro de video, *frame* por *frame*. Visto en continuo, el resultado ofrece una primera visualización de la “mancha” o “nube” de gente presente en nuestro campo visual.

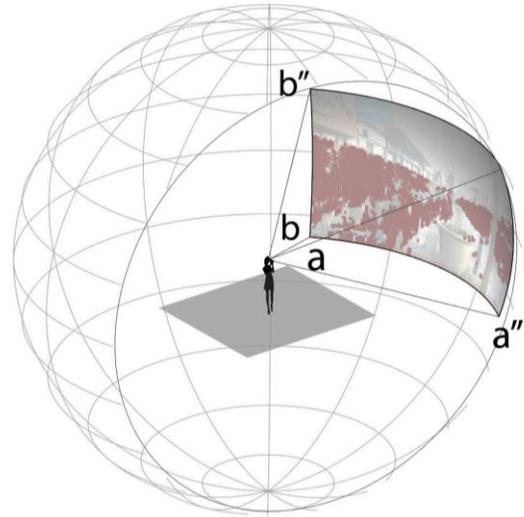


Figura 3: Visualización de parámetros a considerar para definir el dominio del campo visual inscrito en una esfera 3d.

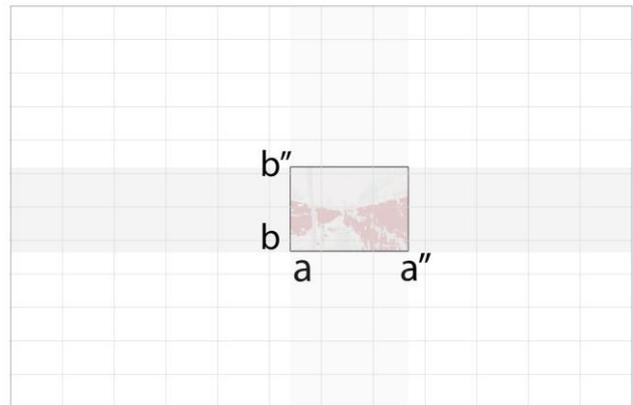


Figura 4: Visualización de parámetros a considerar para definir el dominio del campo visual mapeado en el desarrollo planar de una esfera 3d.

Resultados

El procesamiento de información arroja resultados que tienden a ser consistentes con los resultados de las lecturas realizadas en los pilotos anteriores mas arriba mencionados. Considerando el total del *frame* como 100%, las lecturas de información en forma de personas arrojan un mínimo observado de 1.8% y un máximo de 24% siendo el promedio de las tres muestras de 4.69%. Una revisión posterior del video revela que el máximo observado corresponde a la presencia de personas que se acercan de forma extraordinaria a la cámara, en lo que podríamos denominar como “factor primer plano”. Estas instancias en donde la información en forma de personas pueden llegar a ocupar gran parte del campo visual, sea porque una persona se acerca directamente a la cámara o bien, porque el espacio se encuentra saturado de gente, constituyen situaciones límite que definen el rango máximo posible a encontrar, el que de momento especulamos ronda el 30%. Del mismo modo, no

es difícil predecir que el 0% puede ser detectable en horarios de noche y madrugada.



Figura 5: Muestras de fotogramas expuestos al algoritmo de detección de movimiento basado en Optical Flow.

Por otro lado, la muestra revela que las capturas correspondientes a calles vehiculares tienden a mantenerse semi-pobladas de personas producto de la franja vehicular como puede apreciarse en las muestras b y c. Esto difiere considerablemente cuando el registro es hecho en una calle peatonal como es el caso de la muestra a, la que fue realizada al centro del "paseo peatonal" de la ciudad. Por último, la cuantificación de la información visual en forma de personas abre la posibilidad de cuantificar por contraste la superficie de información en forma de espacio urbano normalmente ocluida durante un trayecto de un peatón promedio por la ciudad.

Por su parte, las curvas obtenidas por muestra son reveladoras en tanto ofrecen una línea del comportamiento de nuestro campo visual en relación a la presencia de gente. Esto dicho, el trabajo se beneficiaría de una interface más directa entre curva y *frame* de modo de poder correlacionar de forma directa la información gráfica con la información visual. Este cruce de información permitiría eventualmente el levantamiento de una base o biblioteca de casos tales como el "factor primer plano".

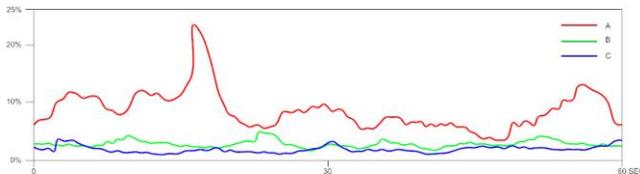


Figura 6: Gráfico % Presencia Humana v/s tiempo, describe la evolución de las muestras A-B-C en el lapsus de 1 minuto.

Tabla 1: Tabla de factores involucrados en el análisis de las muestras A, B, C, sus resultados y sus equivalencias en ejercicios anteriores.

	A	B	C
Latitud	-36.826.081	-36.826.592	-36.827.102
Longitud	-73.049.157	73.050.437	-73.051.686
Ratio A	0.66 To 0.48	0.65 To 0.45	0.67 To 0.49
Ratio b	0.59 To 0.39	0.59 To 0.39	0.59 To 0.39
Op fl %	9,07	3,00	1,99
max	24,40	6,12	5,20
min	3,41	2,97	1,83
oclu/seg	30,20	20,25	18,32
int local	68,30	54,50	49,80
polaroid	4.4	4.4	4.4

Conclusiones

La técnica de análisis de movimiento a través de fotogramas con ayuda de flujo óptico permite capturar, leer y manejar los patrones de presencia y ausencia de gente en el campo visual fotográficamente abstraído de forma sistemática y automatizada. Es decir, permite visibilizar el comportamiento metamórfico de la mancha o nube conformada por personas. Si bien, la técnica de flujo óptico permite una aproximación satisfactoria, esta deja fuera todas aquellas personas que encontrándose presentes en la imagen, no están en movimiento, sea porque están sentadas, sea porque están paradas. Esto sugiere de momento que una lectura tendiente a la exactitud debería usar técnicas de análisis mixtas o bien, de mayor refinamiento técnico. La estandarización del registro obtenido bajo formato universal ofrece la posibilidad de compatibilizar las lecturas de los índices de presencia de personas en el espacio urbano analizado con una diversidad de plataformas tales como aquellas utilizadas para vigilancia y seguridad ciudadana, monitoreo de plusvalía urbana, desarrollo inmobiliario, asesoría de *retail*, etc. Representa además objeto de potencial interés para herramientas consolidadas de análisis urbano tales como las derivadas de la familia de software de Space Syntax. Esto, pues a diferencia del mapa axial, no realiza predicciones relativas a los patrones de presencia y ausencia de gente en el espacio urbano; los revela *ipso facto*. En tanto tal, es una herramienta con el potencial de reducir a 0 la especulación al respecto.



Figura 7: Optical flow acumulado en un minuto como campo visual.

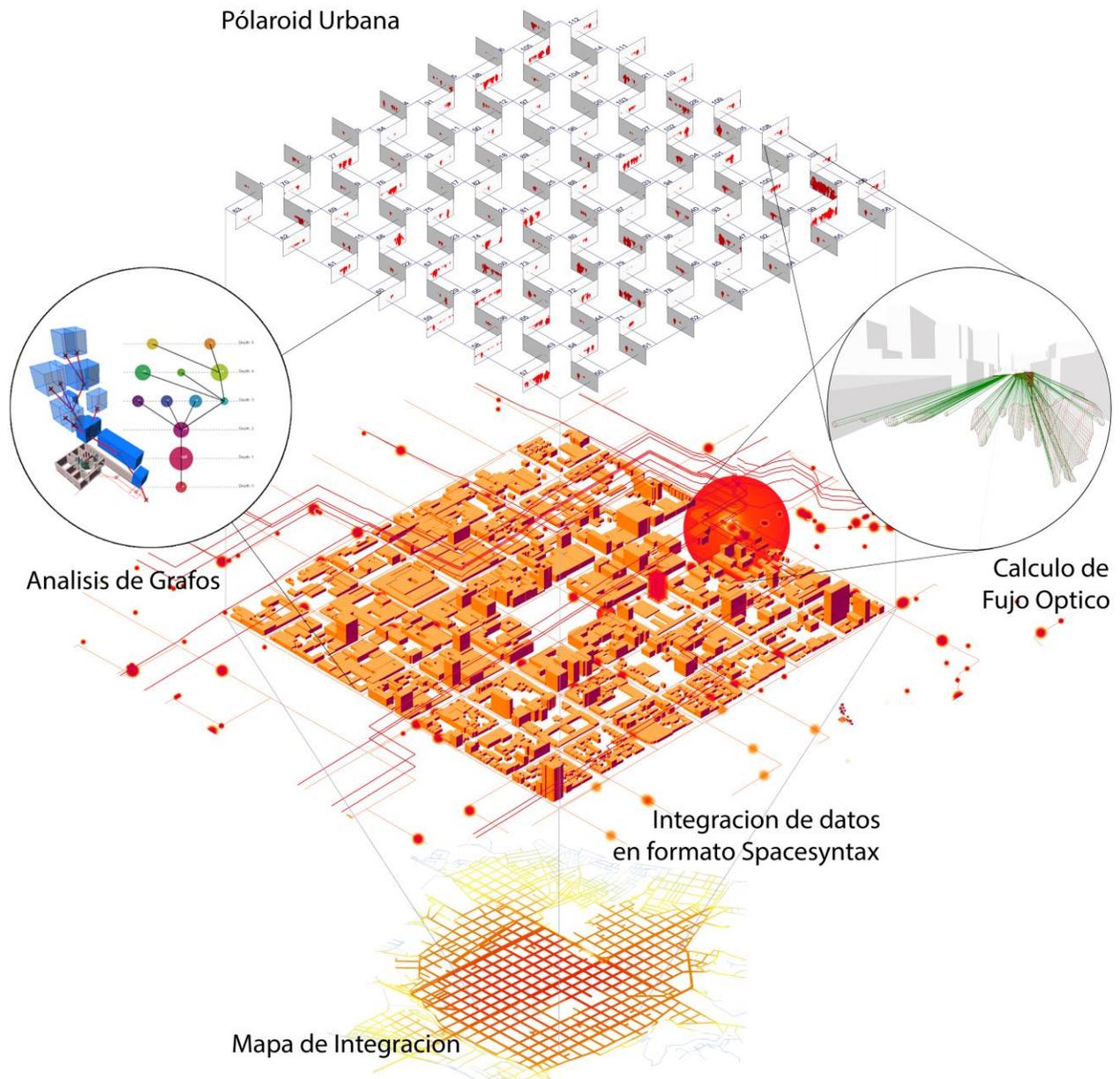


Figura 8: Representación del flujo de datos en el ejercicio: inserción de los datos obtenidos en relación a las experiencias Polaroid Urbana y análisis de integración.

Bibliografía

Araneda, C. (2013a). The Jacobs' Urban Lineage Revisited. Analytical rudiments for the further development of the phenomenological approach to the study of the perception of people in urban space implicit in Jane Jacobs' work, Proceedings of the 31st eCAADe Conference, Delft, Netherlands, Vol.2, pp. 355-363.

Araneda, C. (2013b). Instantánea Urbana. Rudimentos Analíticos para una Aproximación Sistemática al Estudio del Comportamiento de Nuestro Campo Visual en Relación a la Percepción de Personas en el Espacio Urbano. Proceedings XVII Congreso SigraDi, Valparaíso, Chile, pp. 124-128.

Araneda, C. (2014). Polaroid Urbana. Herramienta fenoménica para el levantamiento y manejo sistemático de información en forma de personas. Revista 180, 34, 30-35.

P. Dollár, S. Belongie, and P. Perona. The Fastest Pedestrian Detector in the West. In BMVC, 2010. 2, 3, 7.

Dollár, P., Wojek, C., Schiele, B., Perona, P. (2012). Pedestrian Detection: An Evaluation of the State of the Art. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 34 (4), 743-761.

S. Seer, N. Brändle, C. Ratti, (2012). Kinects and human kinetics: A New Approach for Studying Crowd Behavior (submitted) arXsssigggv: 1210, p. 2838.

Tang, S., Andriluka, M., Schiele, B (2012). Detection and Tracking of Occluded People. British Machine Vision Conference (BMVC). Retrieved from <http://www.bmva.org/bmvc/2012/BMVC/paper009/>

Tang, S., Andriluka, M., Schiele, B, Milan, A., Schindler, K., Roth, S. (2013). Learning People Detectors for Tracking in Crowded Scenes. IEEE International Conference on Computer Vision (ICCV), December, (2013) Retrieved from <https://www.d2.mpi-inf.mpg.de/content/learning-people-detectors-tracking-crowded-scenes>

Tang, S., Andriiuka, M., Schiele, B., Detection and Tracking of Occluded People, S. Tang, M. Andriluka and B. Schiele. (2014). Retrieved from <https://www.d2.mpi-inf.mpg.de/publications>