

Resumen de Tesis Doctoral: Integración escalable de realidad aumentada basada en imágenes y rostros

Programa: Doctorado en Ciencias Informáticas
Facultad de Informática - UNLP
Tesisista: Nahuel Adiel Mangiarua
Director: Jorge Ierache
Codirector: María José Abásolo
Fecha de defensa: 18/12/2020

Grupo de Investigación en Realidad Aumentada Aplicada
Departamento de Ingeniería, UNLaM, Buenos Aires, Argentina
nmangiarua@unlam.edu.ar

Resumen La realidad aumentada permite la adición de elementos digitales sobre la visión natural de las personas. La tesis aquí resumida se enfoca en desarrollar una arquitectura que permita la integración de realidad aumentada basada en imágenes arbitrarias y la realidad aumentada basada en rostros de forma escalable. Se estudiaron los procesos involucrados y los algoritmos necesarios para solventar cuellos de botella. Los resultados obtenidos sugirieron la utilización de múltiples bucles con esquemas de ejecución asíncrono y paralelos así como al algoritmo de búsqueda aproximada de vecinos más cercanos HNSW como el más apropiado para este contexto particular.

Palabras Claves: *Realidad Aumentada; Reconocimiento Facial; Escalabilidad; Arquitectura de Integración; Inferencia biométrica; Approximate Nearest Neighbor Search*

1. Temas de investigación

La Realidad Aumentada (RA) se encuentra inmersa en lo que se ha llamado el continuo de la virtualidad, introducido por (Milgram y Kishino 1994). Este continuo que se extiende desde el ambiente real que nos rodea hasta un ambiente completamente virtual, artificial. La RA consiste en la creación de un entorno en el que la información y los objetos virtuales se fusionan con la realidad, ofreciendo al usuario una experiencia enriquecida, incrementando su percepción natural. La RA puede ser usada para expandir nuestros sentidos, define una visión directa o indirecta de un entorno físico del mundo real, cuyos elementos se combinan con elementos virtuales, como pueden ser textos, imágenes, audio o videos para la creación de una realidad mixta en tiempo real (Abásolo Guerrero et al. 2011).

Una aplicación de RA, en la mayoría de los casos, está conformada mínimamente por un conjunto de cuatro elementos básicos que necesitan estar vinculados de cierta manera para que la aplicación logre su cometido. Los elementos en cuestión son:

- Un elemento que capture las imágenes de la realidad
- Un elemento sobre el que proyectar la integración de las imágenes reales con los contenidos virtuales.
- Una arquitectura de software, embebida o no en un marco de trabajo
- Un disparador o elemento que actúe como desencadenante de RA.

RA basada en imágenes

En particular, la RA basada en el reconocimiento de imágenes planas consiste en utilizar como disparador y anclaje al mundo físico, una imagen plana arbitraria que puede ser detectada, permitiendo la incorporación de información virtual de forma espacial y

contextualmente coherente con la realidad. A diferencia de las imágenes denominadas marcadores, que poseen un patrón geométrico dominante para facilitar su detección y un código visual embebido para diferenciar uno de otro, la imagen arbitraria puede ser entendida como una fotografía, que captura y representa la realidad de un instante congelado en el tiempo. Si bien la realidad física que nos rodea se encuentra en continuo cambio, muchos de estos cambios son lo suficientemente graduales como para considerarlos constantes durante el tiempo de vida útil de una determinada aplicación que utilice esta tecnología.

Escalabilidad y búsqueda de imágenes

La escalabilidad en términos de RA puede naturalmente relacionarse a múltiples aspectos de su ejecución. La acepción utilizada en el contexto particular del trabajo de tesis, hace referencia específicamente a la cantidad de objetivos de aumentación que una arquitectura, u aplicación que la utilice, pueda contener y aumentar. En este sentido no se considera la cantidad de objetivos (imágenes o rostros) que se estén aumentando en simultáneo sobre un mismo cuadro de vídeo o imagen, sino al tamaño del repositorio de objetivos que un software puede reconocer y aumentar individualmente, sin depender de procesamiento externo.

Detección y reconocimiento de rostros e inferencia biométrica

La detección de rostros en tiempo real en un flujo de vídeo, aunque aún un área de investigación activa, resulta posible y aplicable desde los aportes introducidos por el famoso trabajo de Viola y Jones (Viola y Jones 2001). Siguiendo la tendencia general en el área de visión de computador, en los años posteriores muchos otros mecanismos basados en características (usualmente llamadas con el término en inglés, *features*) complejas o esquemas de aprendizaje de máquina simples se han propuesto, no sólo para aumentar el desempeño sino para incrementar el rango de ángulos y de variación lumínica aceptables, como los presentados en (Jegou, Douze, y Schmid 2008; Wang, Han, y Yan 2009; Waring y Liu 2005; Wu y Nevatia 2007). En la actualidad, el problema de la detección de rostros es abordado mayormente mediante la aplicación de técnicas avanzadas de aprendizaje de máquina como presenta (Dantone et al. 2012), en particular destacan las redes neuronales convolucionales y sus arquitecturas derivadas como presentan (Girshick 2015; Li et al. 2015).

Por su parte, la inferencia de parámetros biométricos a partir de imágenes es un campo con numerosas especializaciones. En particular, para en la detección de emociones a través del rostro encontramos un gran número de técnicas tanto a nivel experimental como a nivel productivo. (Corneanu et al. 2016) nos ofrece una extensiva introducción y recorrido por los más importantes aportes hasta la actualidad a la vez que (Valstar et al. 2017) describe los esfuerzos por definir un marco de trabajo y evaluación estandarizado mediante las competencias FERA (Facial Expression Recognition and Analysis).

2. Motivación y objetivos

Tanto la RA basada en imágenes arbitrarias como el reconocimiento facial son utilizados por un número de sistemas, aplicaciones o frameworks en diversos campos de aplicación. Sin embargo, no existen en la actualidad frameworks cuyas arquitecturas integren la capacidad de reconocer imágenes y rostros de manera simultánea y escalable, es decir con un número elevado de objetivos a aumentar. A su vez, ningún framework integra la capacidad de efectuar inferencia biométrica de información a partir de las imágenes percibidas, en particular la de rostros humanos.

Se planteó como objeto principal del trabajo de tesis el diseñar una arquitectura escalable de RA basada en el reconocimiento visual monocular de imágenes y rostros humanos, con capacidad de inferencia de datos biométricos, que no haga uso de servicios externos para su etapa de explotación.

En este orden se plantearon como objetivos particulares:

- Establecer los procesos y sus pasos necesarios para efectuar aumentación de imágenes, detección y reconocimiento de rostros e inferencia de información biométrica.
- Analizar comparativamente la complejidad computacional teórica y la carga de procesamiento empírica de cada paso de los procesos de RA, considerando en particular distintas variaciones de algoritmos disponibles para la búsqueda y descripción de POI.
- Diseñar una arquitectura que integre los procesos descritos, contemplando la ejecución paralela y/o asíncrona, identificando el o los pasos que resulten en el principal cuello de botella con respecto a la escalabilidad.
- Analizar comparativamente la velocidad y precisión de los algoritmos aplicables para aliviar o solventar el o los cuellos de botella detectados.
- Diseñar criterios de evaluación y conjuntos (sets) de datos de prueba para los algoritmos aplicables a los cuellos de botella que sean representativos del dominio de explotación propuesto.
- Incorporar al diseño un mecanismo de integración abierto que facilite el agregado futuro de algoritmos de inferencia biométrica adicionales en la arquitectura propuesta.

Mientras que no se busca competir con sistemas existentes en términos del refinamiento y calidad que estos han logrado con años de desarrollo continuo, se desarrollo una integración de las tecnologías propuestas con una escalabilidad superadora.

3. Aportes y Conclusiones

El trabajo de tesis propone una arquitectura escalable que integra la RA basada en imágenes arbitrarias con la detección y reconocimiento de rostros humanos junto con la inferencia de datos biométricos a partir de ellos.

Partiendo desde la definición de los procesos para RA basada en imágenes y rostros, se determina cuáles son los pasos necesarios para lograr respectivamente y se estudia la complejidad computacional teórica de cada uno junto con la distribución de carga de procesamiento relativa. Estableciendo el algoritmo ORB o la combinación de ORB y FREAK como alternativas viables para la detección y descripción de POI en imágenes, se identifican los pasos de búsqueda de correspondencias entre descriptores, tanto de imágenes (POI) como de rostros como los cuellos de botella de cada proceso. Se selecciona para la descripción de rostros algoritmos de redes neuronales convolucionales entrenadas con el error por tripletas que producen descriptores continuos de 128 dimensiones y se establece que este paso debe ser implementado de forma asíncrona debido a su tiempo de ejecución.

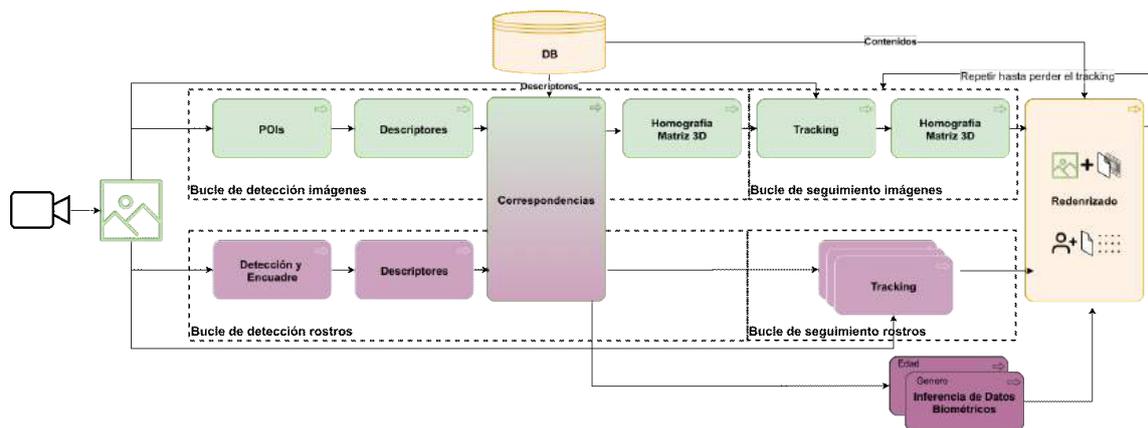
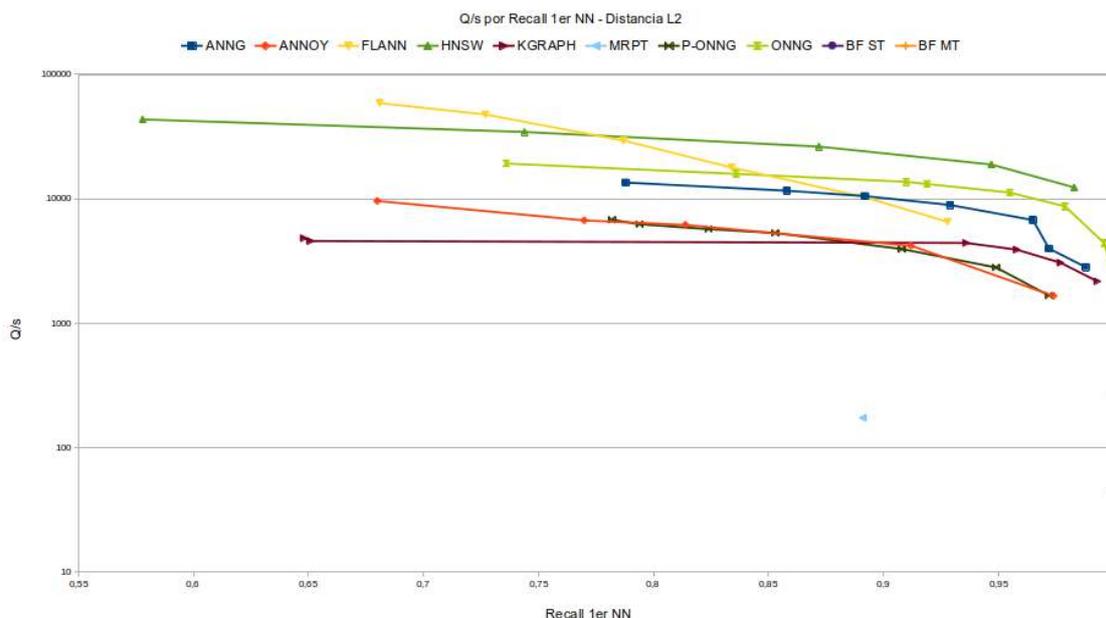


Figura 1: Diagrama conceptual de la arquitectura propuesta. Los recuadros de línea punteada delimitan cada uno de los 4 bucles.

Se presenta el diseño de una arquitectura integrada, compuesta por cuatro bucles en un flujo de ejecución alternante con derivación de tareas asíncronas en un esquema de ejecución paralelo para las tareas de descripción de rostros e inferencia de información biométrica, como se aprecia en la figura 1. Los algoritmos seleccionados permiten considerar el cuello de botella de cada proceso integrado como un mismo problema el cual se propone aliviar con el uso de algoritmos de búsqueda ANN.

Para lograr la escalabilidad de la arquitectura integrada, se realizan una serie de experimentos para analizar comparativamente la velocidad, precisión y estabilidad de distintos algoritmos de búsqueda ANN, estableciendo un marco de evaluación y conjuntos de datos específicos para el contexto de la RA basada en imágenes y reconocimiento de rostros. En la figura 2 se observa particularmente la medida de queries (operaciones de búsqueda) por segundo de cada algoritmo de acuerdo a su precisión (recall). Se establece y valida la superioridad del algoritmo HNSW para la tarea en este contexto particular, difiriendo de los resultados provistos por otros autores para contextos generales.



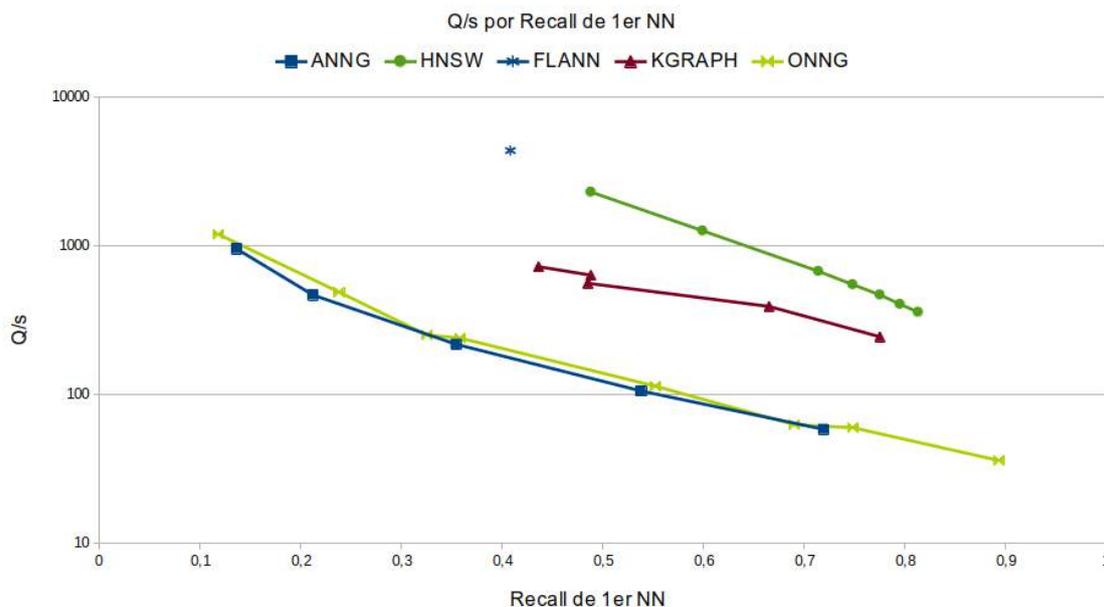


Figura 2: Gráficos de cantidad de operaciones por segundo según precisión (recall) para múltiples algoritmos de búsqueda ANN probados. Mientras más arriba a la derecha mejor es el algoritmo.

También se obtiene como aporte secundario un nuevo esquema de evaluación de algoritmos de búsqueda de vecinos más cercanos aproximados específico para el contexto de la RA. El mismo, implica la evaluación de esta familia de algoritmos utilizando sets de datos donde la variación entre los elementos query y los elementos de entrenamiento, base u originales sea reducida. Se establece un porcentaje de variación de entre el 5% y el 15% de acuerdo al tipo función de distancia utilizada pero se recomienda ajustar estos valores según el algoritmo específico que genera los elementos. Con estos sets de datos se propone un esquema donde los algoritmos de búsqueda de ANN sean comparados por su recall en los dos primeros vecinos.

Se crea un prototipo demostrador experimental que implementa la arquitectura propuesta en C++, compatible con cualquier plataforma para la que se lo compile. La misma optimiza la utilización de múltiples hilos de procesamiento y permite la posibilidad de integración con sistemas de más alto nivel para la generación y explotación de aplicaciones de RA. En la misma se implementa una interfaz abstracta que permite la incorporación dinámica de algoritmos de inferencia biométrica que serán automáticamente ejecutados de manera paralela y asíncrona.

Contexto

La tesis para el doctorado en Ciencias Informáticas de la Facultad de Informática de la UNLP se radicó en el Grupo de Realidad Aumentada Aplicada (GRAA) del Departamento de Ingeniería de la UNLaM, fue desarrollada en el marco del proyecto PROINCE C-231. La dirección estuvo a cargo de Jorge Ierache (GRAA, UNLaM) y la co-dirección de Maria Jose Abasolo (LIDI, UNLP).

Futuros trabajos

Si bien el marco de evaluación de algoritmos de ANN para el contexto de la RA propuesto predice con mayor precisión el desempeño de los algoritmos, se requiere continuar sobre esta

línea para establecer condiciones aún más específicas y con el fin de desarrollar una métrica de comparación concreta más precisa.

Se plantea abordar la compilación cruzada del prototipo de implementación a plataformas ARM, así como la integración con sistemas de alto nivel, específicamente el motor de gráficos Unity3D.

También se plantea la integración continua de algoritmos de inferencia de datos biométricos a medida que sean desarrollados y publicados por sus respectivos autores.

Finalmente se planea, continuando con la línea de investigación bajo el marco del proyecto PROINCE C-231 Comandos de Voz y Reconocimiento Facial para Aplicaciones de Realidad Aumentada, el desarrollo de un prototipo demostrador en el contexto de la emergentología.

Publicaciones relacionadas con la tesis

Revistas, series internacionales y capítulos de libro

Mangiarua, Nahuel, Jorge Ierache, and María José Abasolo. "Scalable Integration of Image and Face Based Augmented Reality." In *Augmented Reality, Virtual Reality, and Computer Graphics*, edited by Lucio Tommaso De Paolis and Patrick Bourdot. Cham: Springer International Publishing, https://doi.org/10.1007/978-3-030-58465-8_18

En esta publicación se comunican los principales elementos de diseño y la arquitectura integrada objeto de este trabajo de tesis.

Mangiarua, Nahuel, Jorge Ierache, and María José Abasolo. "Implementation of an Open Source Based Augmented Reality Engine for Cloud Authoring Frameworks." *Journal of Computer Science and Technology* 19, no. 2 (October 2019): e16. <https://doi.org/10.24215/16666038.19.e16>.

En esta publicación se comunica el diseño del proceso basado en imágenes y su análisis particular así como consideraciones para el aprovechamiento de capacidades de la nube para herramientas de autor.

Mangiarua, Nahuel, Jorge Ierache, Martín Becerra, Hernán Maurice, Santiago Igarza, and Osvaldo Sposito. "Templates Framework for the Augmented Catalog System." In *Computer Science – CACIC 2018*, edited by Patricia Pesado and Claudio Aciti, 267–276. Cham Springer International Publishing: Springer International Publishing, 2019. <http://sedici.unlp.edu.ar/handle/10915/73478>.

En esta publicación se comunican los principales elementos de diseño y la implementación del concepto de template para sistemas de autor que facilita el trabajo a usuarios finales frente a un incremento de escala en el sistema.

Ierache, Jorge, Nahuel Mangiarua, Martín Ezequiel Becerra, and Santiago Igarza. "Framework for the Development of Augmented Reality Applications Applied to Education Games." In *Proceedings of the 5th International Conference on Augmented Reality, Virtual Reality, and Computer Graphics*, edited by Lucio Tommaso De Paolis and Patrick Bourdot, 340–350. Cham Springer International Publishing: Springer International Publishing, 2018. https://doi.org/10.1007/978-3-319-95270-3_28.

Esta publicación complementaria comunica un framework especializado desarrollado sobre el Sistema de Catálogos Virtuales Aumentados, el cual constituye un caso de aplicación potencial para la implementación de la arquitectura propuesta.

Congresos internacionales

Montalvo, C., F. Petrolo, D. Sanz, Nahuel Mangiarua, N. Verdicchio, S. Igarza, and J. Ierache. “Knowledge Based Augmented Card System for Medical Assistance Over Mobile Devices.” In Selected Papers of the XXI Argentine Congress of Computer Science, 257–65, 2017.

Ierache, Jorge, Nicolás Nazareno Verdicchio, N. Duarte, and Nahuel Mangiarua. “Augmented Reality Card System for Emergency Medical Services.” In Proceedings of the International Work-Conference on Bioinformatics and Biomedical Engineering, 487–94. Granada, Spain, 2016.

Ierache, Jorge, Nahuel Mangiarua, N. Verdicchio, and D. Sanz. “Augmented Card System Based on Knowledge for Medical Emergency Assistance.” In 2016 IEEE Congreso Argentino de Ciencias de la Informática y Desarrollos de Investigación (CACIDI), 1–3, 2016. <https://doi.org/10.1109/CACIDI.2016.7785979>.

Ierache, Jorge, Nahuel Mangiarua, S. Bevacqua, N. Verdicchio, and M. Becerra. “Development of a Catalogs System for Augmented Reality Applications.” In World Academy of Science, Engineering and Technology, International Science Index, 9:1–7, 2015.
<http://waset.org/Publications/development-of-a-catalogs-system-for-augmentedreality-applications/10000077>.

Congresos nacionales y workshops

Mangiarua, Nahuel, Jorge Ierache, and María José Abásolo Guerrero. “Avances en Línea de Investigación Doctoral: Integración Escalable de Realidad Aumentada Basada en Imágenes y Rostros.” In Libro de actas del XXII Workshop de Investigadores en Ciencias de la Computación. Santa Cruz, Argentina, 2020.

Mangiarua, Nahuel, Jorge Ierache, Martín Ezequiel Becerra, Hernán Maurice, Santiago Igarza, and Osvaldo Mario Sposito. “Framework para la generación de templates en sistema de catálogos de RA.” In Libro de actas del XXIV Congreso Argentino de Ciencias de la Computación. La Plata, Argentina, 2018.
<http://sedici.unlp.edu.ar/handle/10915/73478>.

Ierache, Jorge, Javier Urien, Nahuel Mangiarua, Martín Ezequiel Becerra, Paola Pezoimburg, Emidio Bueno, Hernán Maurice, and Eric S. Auchterberge. “Desarrollo de un prototipo de RA de tipo geoespacial.” In Libro de actas del XII Jornadas de Vinculación Universidad-Industria, 5–8, 2018.
<http://sedici.unlp.edu.ar/handle/10915/71289>.

Becerra, Martín Ezequiel, Nahuel Mangiarua, Santiago Igarza, Jorge Ierache, and María José Abásolo Guerrero. “Líneas de investigación del grupo de RA aplicada: templates de catálogos aumentados integración escalable de RA basada en imágenes y rostros aumentación de sistemas SCADA en el contexto de la industria 4.0.” In Libro de actas del XX Workshop de Investigadores en Ciencias de la Computación, 2018.
<http://sedici.unlp.edu.ar/handle/10915/67455>.

Mangiarua, Nahuel, Cristian Montalvo, Facundo Petrolo, Diego Rubén Sanz, Nicolás Nazareno Verdicchio, E. Lobatto, A. Rosenthal, Martín Ezequiel Becerra, Santiago Igarza, and Jorge Ierache. “Framework para la generación de templates en sistemas de catálogos de RA.” In Libro de actas del XIX Workshop de Investigadores en Ciencias de la Computación, 2017. <http://sedici.unlp.edu.ar/handle/10915/61824>.

Verdicchio, Nicolás Nazareno, Diego Rubén Sanz, Cristian Montalvo, Facundo Petrolo, Nahuel Mangiarua, Santiago Igarza, and Jorge Ierache. “Sistema de catálogo virtual aumentado: integración de framework especializado orientado a juegos didácticos.” In Libro de actas del XI Congreso de Tecnología en Educación y Educación en Tecnología, 2016. <http://hdl.handle.net/10915/54651>.

Trabajos previos

Bevacqua, Sebastián Ariel, Santiago Igarza, Nahuel Mangiarua, Martín Ezequiel Becerra, Nicolás Nazareno Verdicchio, Fernando Martín Ortiz, Diego Rubén Sanz, Nicolás Daniel Duarte, Matías Ezequiel Sena, and Jorge Ierache. “Líneas de investigación del grupo de Realidad Aumentada Aplicada de UNLaM.” In Libro de actas del XVII Workshop de Investigadores en Ciencias de la Computación. Salta, Argentina, 2015. <http://sedici.unlp.edu.ar/handle/10915/45657>.

Mangiarua, Nahuel, Jorge Ierache, and Martin Becerra. “Herramienta de Realidad Aumentada Para La Explotación de Material Didáctico Tradicional.” In Libro de Actas Del IX Congreso de Tecnología En Educación En Tecnología, 240–54. Chilecito, Argentina, 2014.

Ierache, Jorge, Santiago Igarza, Nahuel Mangiarua, Martín Ezequiel Becerra, Sebastián Ariel Bevacqua, Nicolás Nazareno Verdicchio, Fernando Martín Ortiz, Diego Rubén Sanz, Nicolás Daniel Duarte, and Esteban de la Llave. “Realidad Aumentada (RA) en el contexto de usuarios finales.” In Libro de actas del XVI Workshop de Investigadores en Ciencias de la Computación, 2014. <http://sedici.unlp.edu.ar/handle/10915/41253>.

Ierache, Jorge, Nahuel Mangiarua, and Sebastián Ariel Bevacqua. “Sistema de catálogo para la asistencia a la creación, publicación, gestión y explotación de contenidos multimedia y aplicaciones de RA.” In Libro de actas del XX Congreso Argentino de Ciencias de la Computación. Buenos Aires, Argentina, 2014. <http://sedici.unlp.edu.ar/handle/10915/42339>.

Ierache, Jorge, Nahuel Mangiarua, and Martin Becerra. “Herramienta de Realidad Aumentada Para Facilitar La Enseñanza En Contextos Educativos Mediante El Uso de Las TICs,” *Revista Latinoamericana de Ingeniería de Software*, 1 (2014): 1–3.

Esta tesis doctoral

Nahuel A. Mangiarua. “Integración Escalable de Realidad Aumentada Basada en Imágenes y Rostros”. Tesis doctoral, Universidad Nacional de La Plata, La Plata, Argentina, 2020. <http://sedici.unlp.edu.ar/handle/10915/111301>

Bibliografía

Abásolo Guerrero, María José, Cristina Manresa Yee, Ramón Más Sansó, y Marcelo Vénere. 2011. *Realidad virtual y realidad aumentada*. Editorial de la Universidad Nacional de La Plata (EDULP). <http://sedici.unlp.edu.ar/handle/10915/18399>.

Corneanu, C.A., M.O. Simón, J.F. Cohn, y S.E. Guerrero. 2016. «Survey on RGB, 3D, Thermal, and Multimodal Approaches for Facial Expression Recognition: History, Trends, and Affect-Related Applications». *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38. <https://doi.org/10.1109/TPAMI.2016.2515606>.

Dantone, M., J. Gall, G. Fanelli, y L. Van Gool. 2012. «Real-Time Facial Feature Detection Using Conditional Regression Forests». En *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2578–85. <https://doi.org/10.1109/CVPR.2012.6247976>.

Girshick, Ross B. 2015. «Fast R-CNN». *CoRR Abs/1504.08083*. <http://arxiv.org/abs/1504.08083>.

Jegou, Herve, Matthijs Douze, y Cordelia Schmid. 2008. «Hamming Embedding and Weak Geometric Consistency for Large Scale Image Search». En *Computer Vision – ECCV 2008: 10th European Conference on Computer Vision*, editado por David Forsyth, Philip Torr, y Andrew Zisserman, 304–317. Marseille, France: Springer. https://doi.org/10.1007/978-3-540-88682-2_24.

Li, H., Z. Lin, X. Shen, J. Brandt, y G. Hua. 2015. «A Convolutional Neural Network Cascade for Face Detection». En *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5325–34. <https://doi.org/10.1109/CVPR.2015.7299170>.

Milgram, Paul, y Fumio Kishino. 1994. «A Taxonomy of Mixed Reality Visual Displays». *IEICE Trans. Information Systems* E77-D, n.º 12: 1321-29.

Valstar, Michel F., Enrique Sánchez-Lozano, Jeffrey F. Cohn, László A. Jeni, Jeffrey M. Girard, Zheng Zhang, Lijun Yin, y Maja Pantic. 2017. «FERA 2017 - Addressing Head Pose in the Third Facial Expression Recognition and Analysis Challenge». *CoRR Abs/1702.04174*. <http://arxiv.org/abs/1702.04174>.

Viola, P., y M. Jones. 2001. «Rapid Object Detection Using a Boosted Cascade of Simple Features». En *1:1-511-1-518*. IEEE Comput. Soc. <https://doi.org/10.1109/CVPR.2001.990517>.

Wang, X., T.X. Han, y S. Yan. 2009. «An HOG-LBP Human Detector with Partial Occlusion Handling». En *2009 IEEE 12th International Conference on Computer Vision*, 32–39. <https://doi.org/10.1109/ICCV.2009.5459207>.

Waring, C.A., y Xiuwen Liu. 2005. «Face Detection Using Spectral Histograms and SVMs». *IEEE Transactions on Systems*. <https://doi.org/10.1109/TSMCB.2005.846655>.

Wu, Bo, y Ram Nevatia. 2007. «Detection and Tracking of Multiple, Partially Occluded Humans by Bayesian Combination of Edgelet Based Part Detectors». *International Journal of Computer Vision* 75 (2). <https://doi.org/10.1007/s11263-006-0027-7>.