Grupo de Ingeniería Electrónica
aplicada a Espacios INteligentes y TRAnsporte

Modelado de arrays de micrófonos como cámaras de perspectiva en aplicaciones de localización de locutores

geintra

Alejandro Legrá-Ríos, Javier Macías-Guarasa, Daniel Pizarro y Marta Marrón-Romera

Departamento de Electrónica – Universidad de Alcalá

email: macias@depeca.uah.es

SAAEI'2011 – Sesión Especial Aplicaciones en espacios inteligentes – 5/julio/2011

Índice

- Introducción
- Descripción del sistema
  - Propuesta general
  - Generación de imágenes de potencia acústica
  - Búsqueda de máximos y cálculo de posición 3D
- Evaluación experimental:
  - Bases de datos y métricas de evaluación
  - Resultados
- Conclusiones y líneas futuras

Introducción

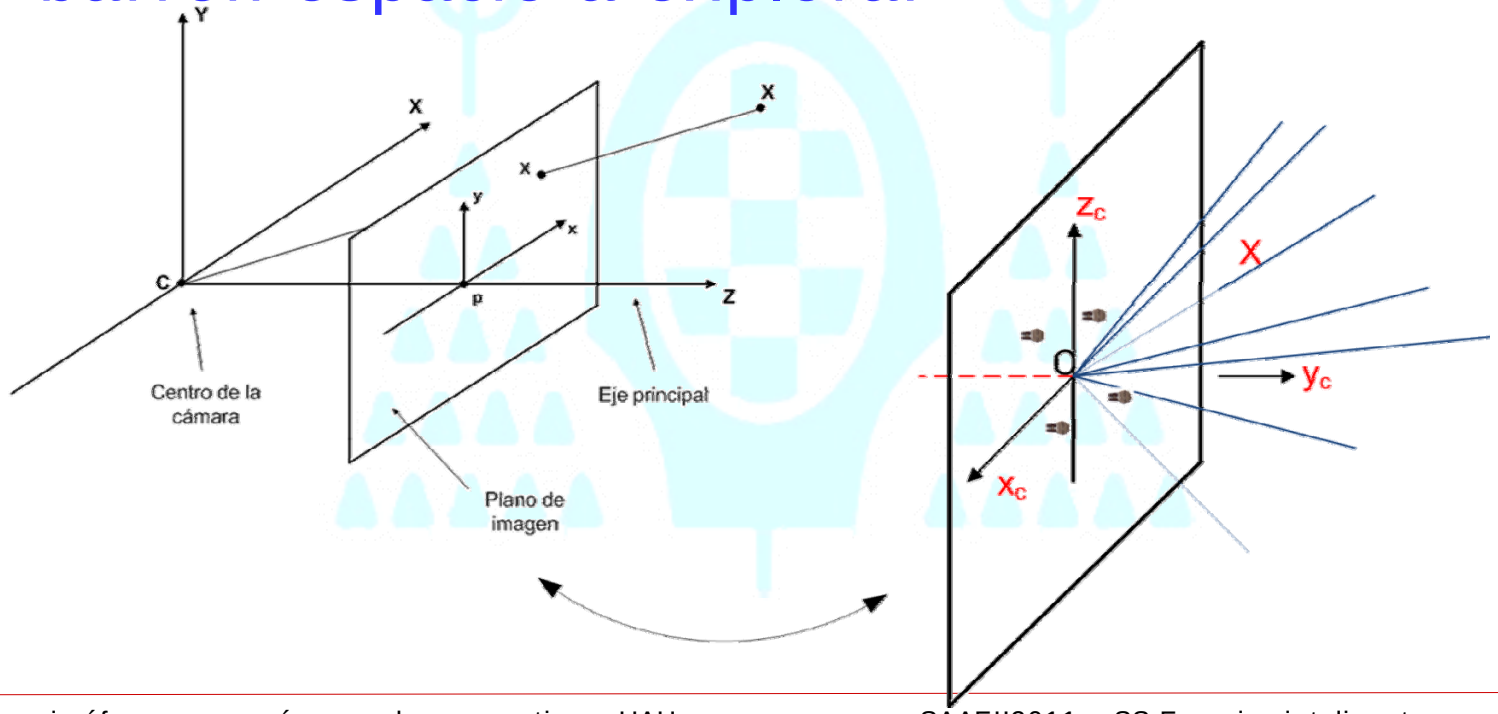
- Contexto:
  - Análisis automático de espacios inteligentes
  - Detección, **localización** y seguimiento de ***locutores***
- Estrategias:
  - Sensores de visión, acústicos, etc.
  - Fusión sensorial
- Nuestra propuesta:
  - Sensado acústico: modelado de *arrays* de micrófonos como cámaras de perspectiva
  - Literatura: tareas distintas y falta evaluación objetiva con datos reales

Evaluación preliminar

Descripción del sistema

Propuestas general

- Cámara:
  - Píxel como intersección rayo con plano imagen
- Array de micrófonos:
  - Rayos *acústicos* parten de centro del array y barren espacio a explorar



Descripción del sistema

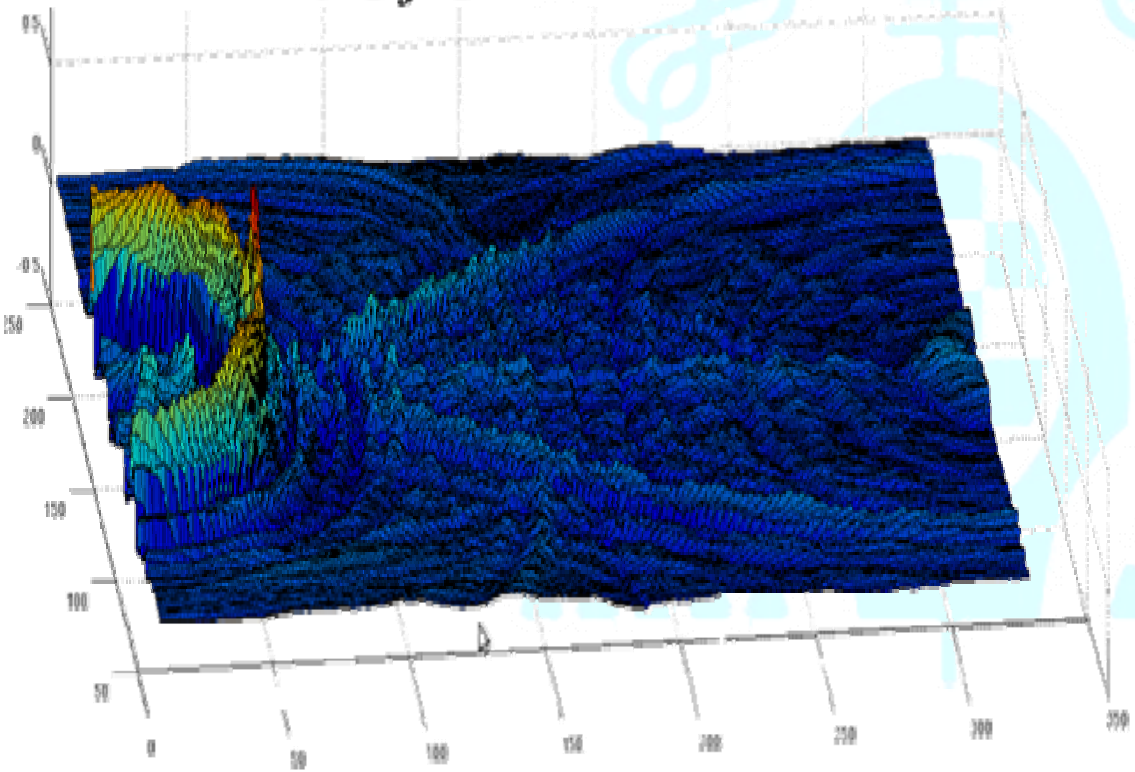
Generación de imágenes *acústicas*

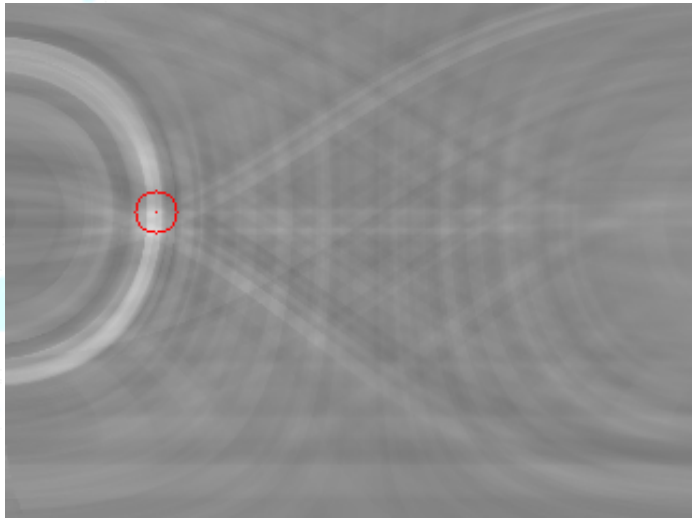
- Generación de mapas de potencia acústica:

$$P_{(\text{azimut,elevación})} = \frac{\sum_{k=1}^{k=n_{(\text{azimut,elevación})}} P(\mathbf{q}_k)}{n_{(\text{azimut,elevación})}}$$

$$P(\mathbf{q}) = \sum_{i=1}^{M}\sum_{j=1}^{M}\int_{-\infty}^{\infty}\Phi_{ij}(\omega)X_i(\omega)X_j'(\omega)e^{j\omega(\Delta_j-\Delta_i)}d\omega$$

SRP-PHAT

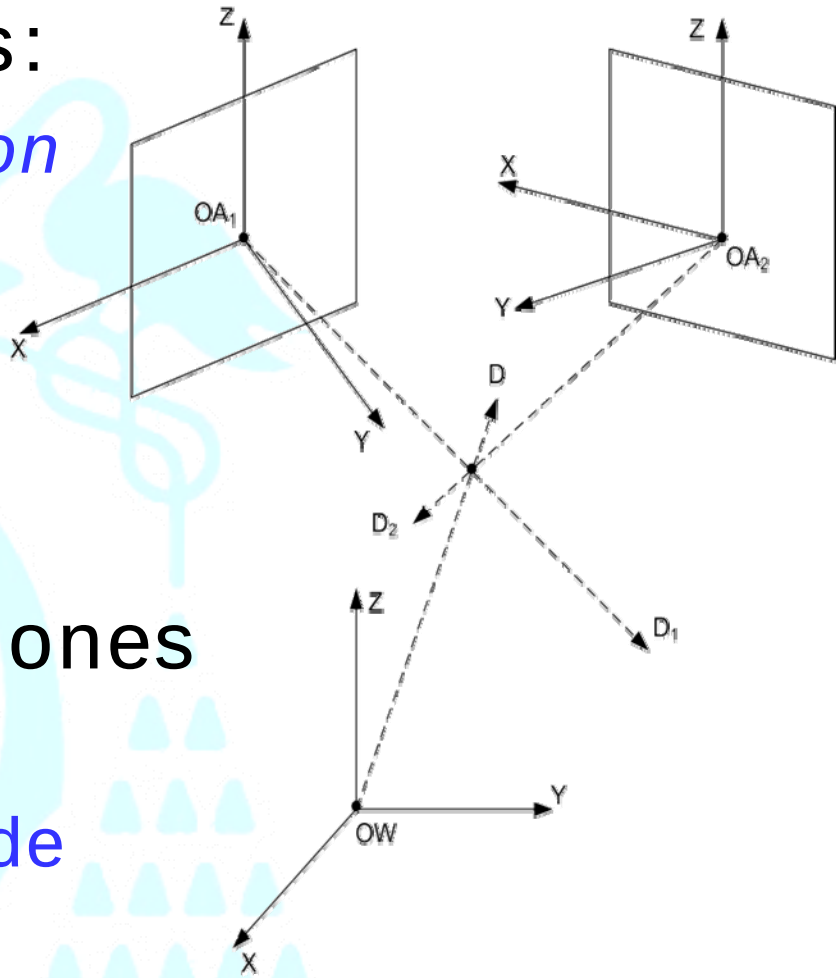




Descripción del sistema

Búsqueda de máximos y cálculo posición 3D

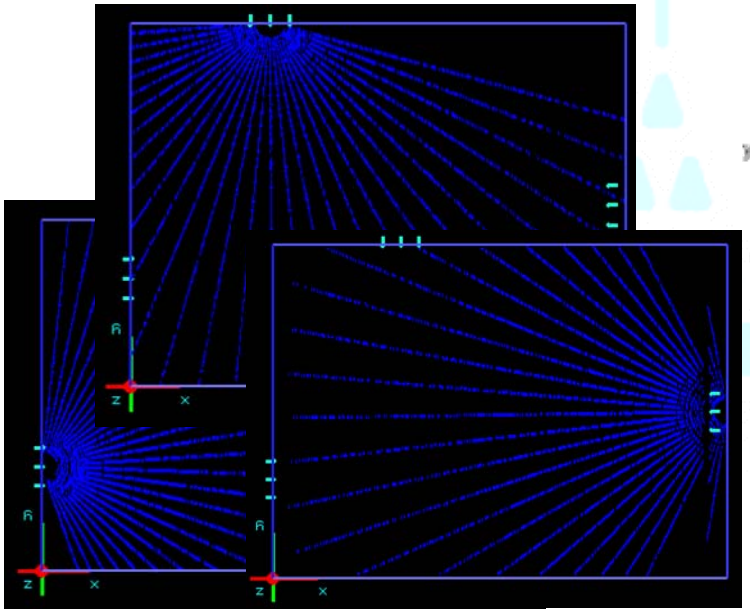
- Estimación de máximos:
  - *Non maximum supression* con aproximación subpixélica
- Cálculo de posición 3D:
  - Triangulación lineal DLT
- Eliminación de estimaciones incoherentes:
  - Para entornos con más de 4 arrays
  - Variante de la técnica *Random Sample Consensus* (RANSAC)

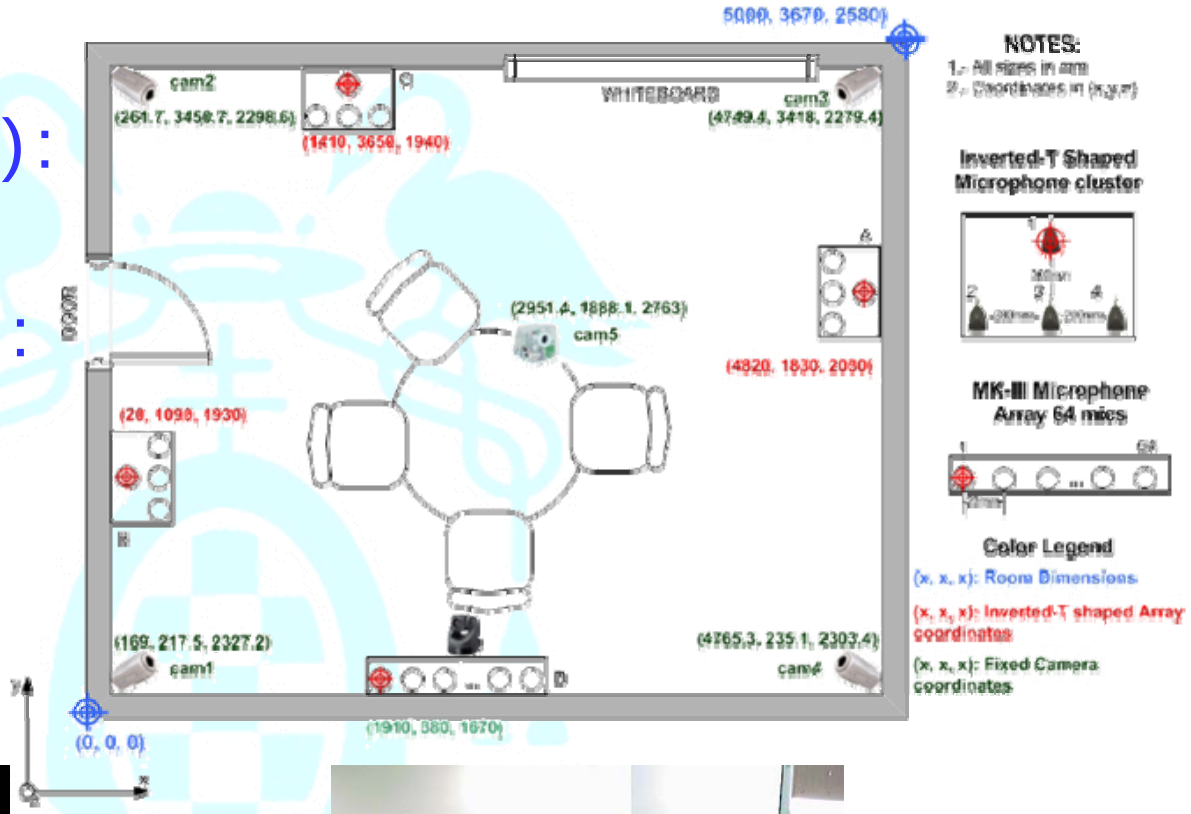


Evaluación experimental

Bases de datos proy. CHIL + eval. CLEAR 2007

- Seminarios:
  - AIT (~40 min.):
    - 3 arrays
  - ITC (~60 min):
    - 6 arrays





Evaluación experimental

Métricas

- Pcor:
  - % tramas con error de posicionamiento inferior a 50cm (*fine error*)
- *AEE fine+gross*:
  - El error promedio ***total*** cometido en todas las estimaciones realizadas
- *AEE fine*:
  - El error promedio en las estimaciones definidas como *fine errors*
- *Deletions*:
  - % de tramas con locutor activo no detectado por el sistema

Evaluación experimental

Resultados

- Variación en resolución imágenes (AIT):
  - Con separación radial en rayo de 100mm

Métricas de CHIL	320 × 240	240 × 180	160 × 120
Pcor	55.0 ± 2.7%	56.0 ± 2.7%	53.0 ± 2.7%
Bias fine (x:y:z)	1 : 29 : 118	1 : 38 : 119	4 : 37 : 118
Bias fine+gross (x:y:z)	23 : 59 : 171	37 : 58 : 175	83 : 92 : 173
AEE fine	246 mm	253 mm	253 mm
AEE fine+gross	634 mm	637 mm	654 mm
Deletion rate	0%	0%	0%
Veces tiempo real	3.17	1.75	0.84

  - SRP-PHAT exhaustivo (tamaño celda 50mm):
    - *Pcor* = 63%
    - *AEE Fine* = 233mm
    - *AEE fine+gross* = 572mm

Resultados prometedores

Evaluación experimental

Resultados

- Variación en separación radial (AIT):

Métricas de CHIL	320×240 Δr 50 mm	320×240 Δr 100 mm	320×240 Δr 200 mm
Pcor	$55.0 \pm 2.7\%$	$55.0 \pm 2.7\%$	$56.0 \pm 2.7\%$
Bias fine (x:y:z)	4 : 29 : 118	1 : 29 : 118	4 : 38 : 113
Bias fine+gross (x:y:z)	27 : 57 : 170	23 : 59 : 171	17 : 60 : 167
AEE fine	247 mm	246 mm	249 mm
AEE fine+gross	631 mm	634 mm	625 mm
Deletion rate	0 %	0 %	0 %
Veces tiempo real	5.23	3.17	1.93

- SRP-PHAT exhaustivo (tamaño celda 50mm):
  - *Pcor* = 63%
  - *AEE Fine* = 233mm
  - *AEE fine+gross* = 572mm

Necesidad de afinar estrategia de separación radial

Evaluación experimental

Resultados

- Estimación de coherencia (ITC):

Métricas de CHIL	Coherencia	Sin Coherencia
Pcor	$60.0 \pm 3.5\,\%$	$12.0 \pm 2.0\,\%$
Bias fine (x:y:z)	151 : 5 : 24	139 : 210 : 82
Bias fine+gross (x:y:z)	351 : 306 : 14	352 : 787 : 126
Bias AEE fine	198 mm	348 mm
Bias fine+gross	759 mm	1020 mm
Deletion rate	25 %	0 %

- SRP-PHAT exhaustivo (tamaño celda 50mm):
  - *Pcor* = 62%
  - *AEE Fine* = 175mm
  - *AEE fine+gross* = 762mm

Clave, pero necesidad de mejorar tasa de borrados

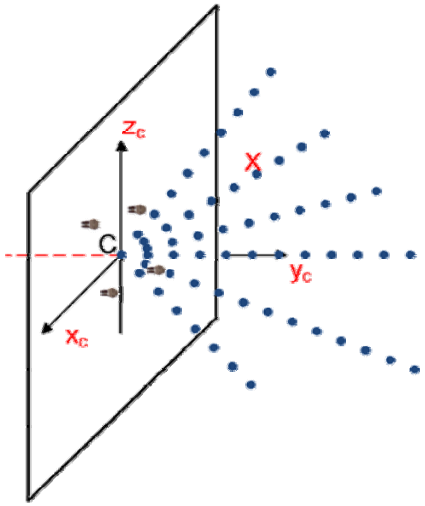
Conclusiones y líneas futuras

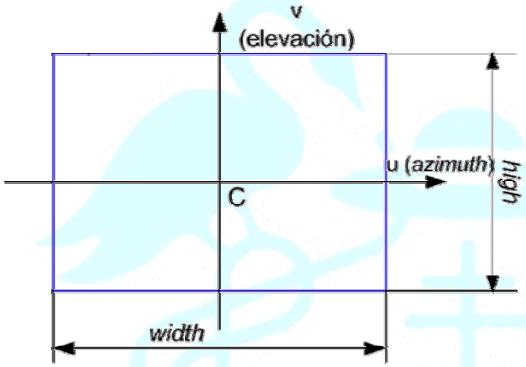
- Propuesta de sistema de localización de locutores:
  - Información acústica
  - Algorítmica de visión, modelando arrays de micrófonos como cámaras de perspectiva
- Resultados:
  - No mejoran algoritmos clásicos
  - Trabajo preliminar: prometedores
- Trabajo futuro:
  - Evaluación exhaustiva
  - Técnicas más elaboradas de estimación de máximos y evaluación de coherencia
  - Incorporación de algoritmos de seguimiento

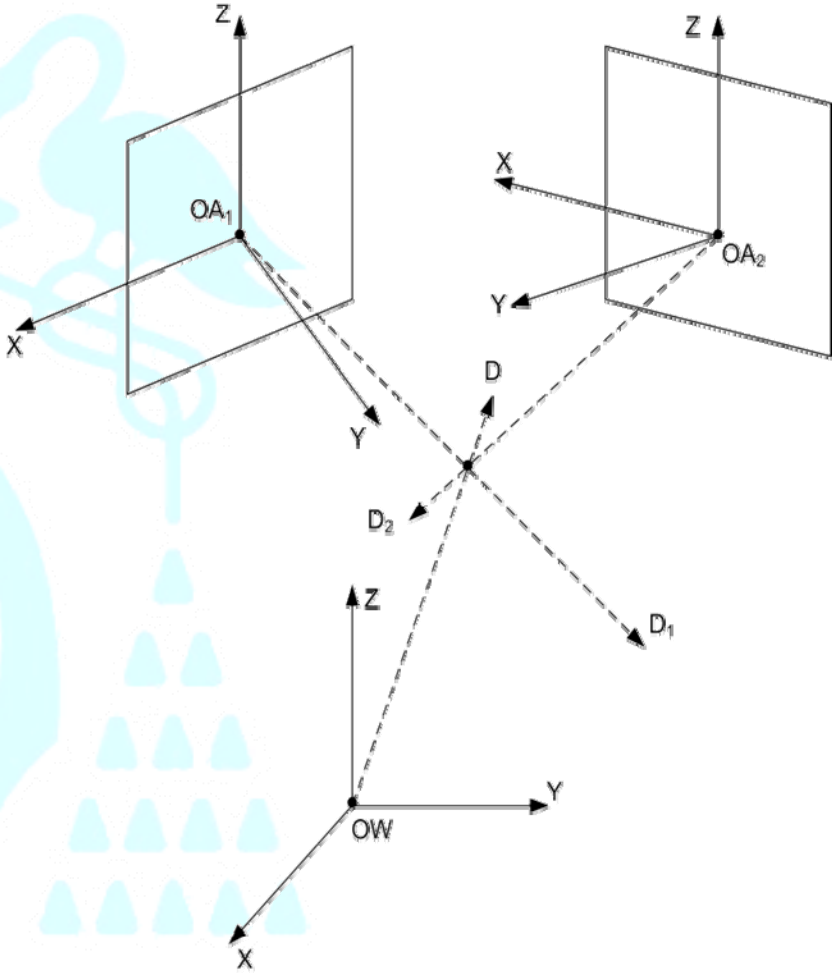
¿ Preguntas ?

Descripción del sistema

Búsqueda de máximos y cálculo posición 3D







$$D_i \times X^{OA_i} = D_i \times (R^{OA_i}[X^{OW} - T^{OA_i}]) = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$A_i X^{OW} = B_i \qquad A_i \in \mathbb{R}^{3\times 3} \qquad B_i \in \mathbb{R}^3$$

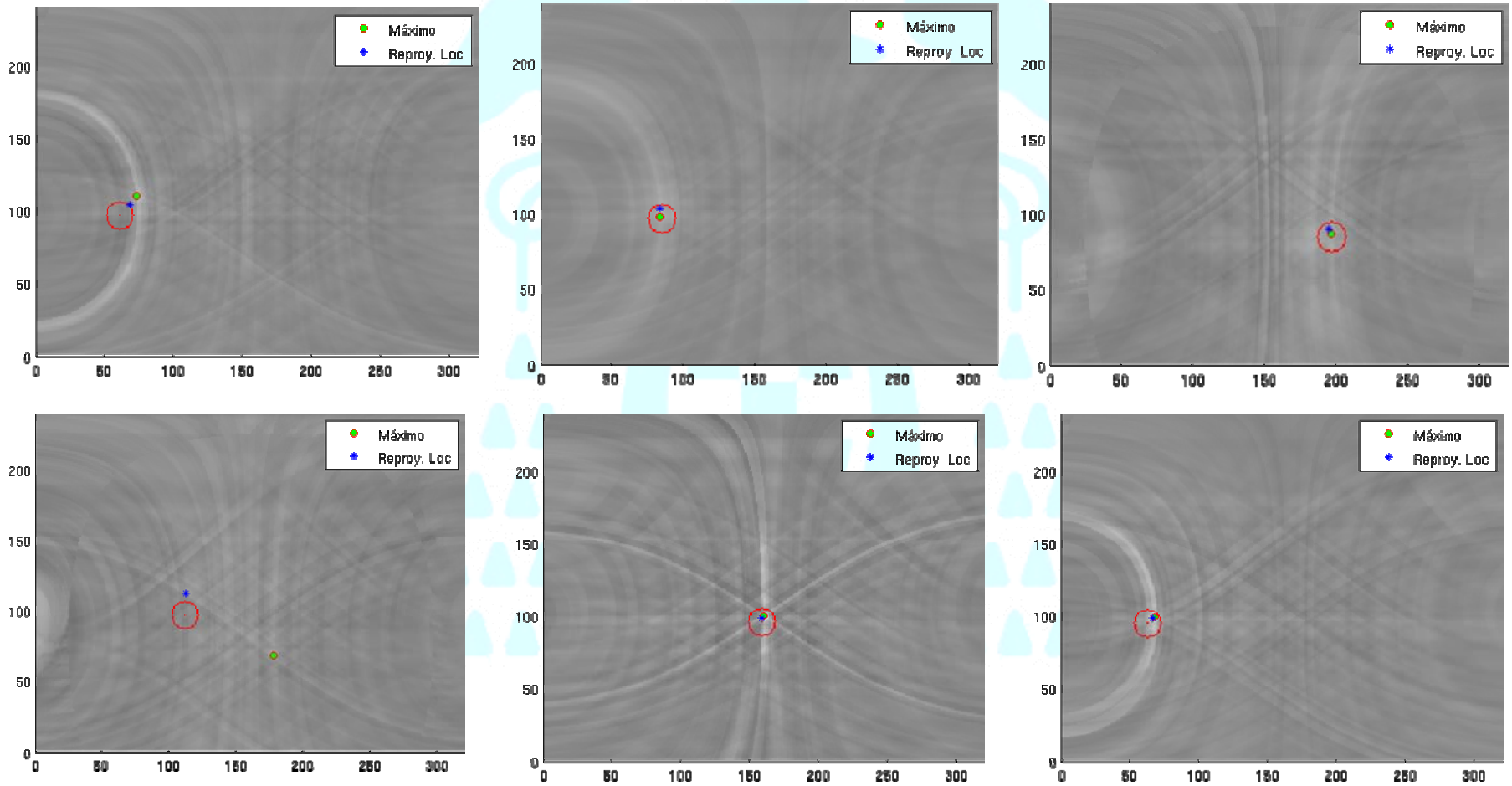
$$AX^{OW} = B \qquad A \in \mathbb{R}^{2N\times 3} \qquad B \in \mathbb{R}^{2N}$$

$$X^{OW} = (A^T A)^{-1} A^T B$$

Evaluación experimental

Resultados

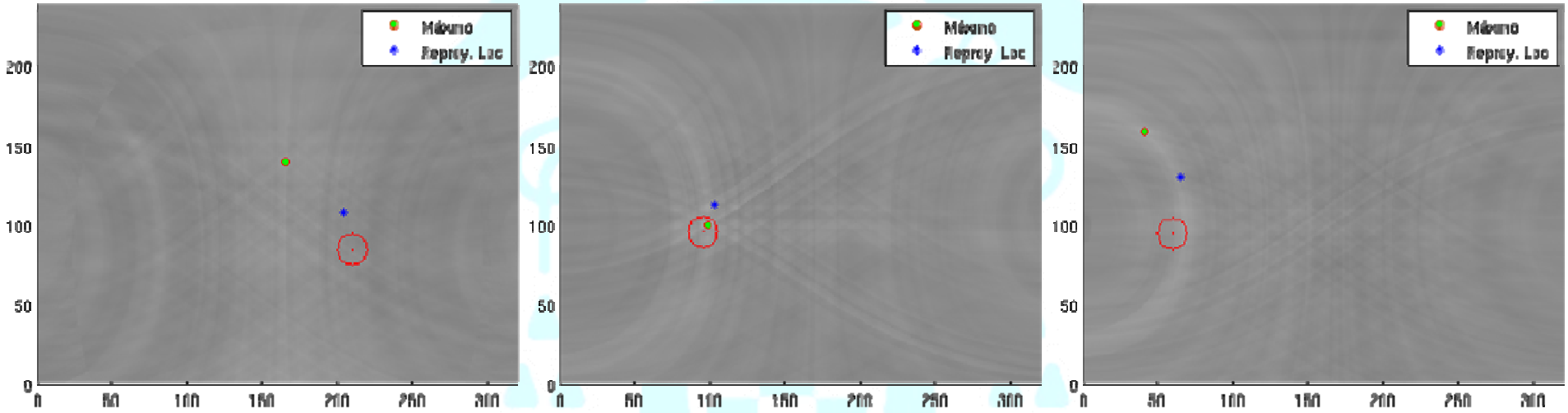
- Experimento base:
  - Error pequeño en estimación DLT



Evaluación experimental

Resultados

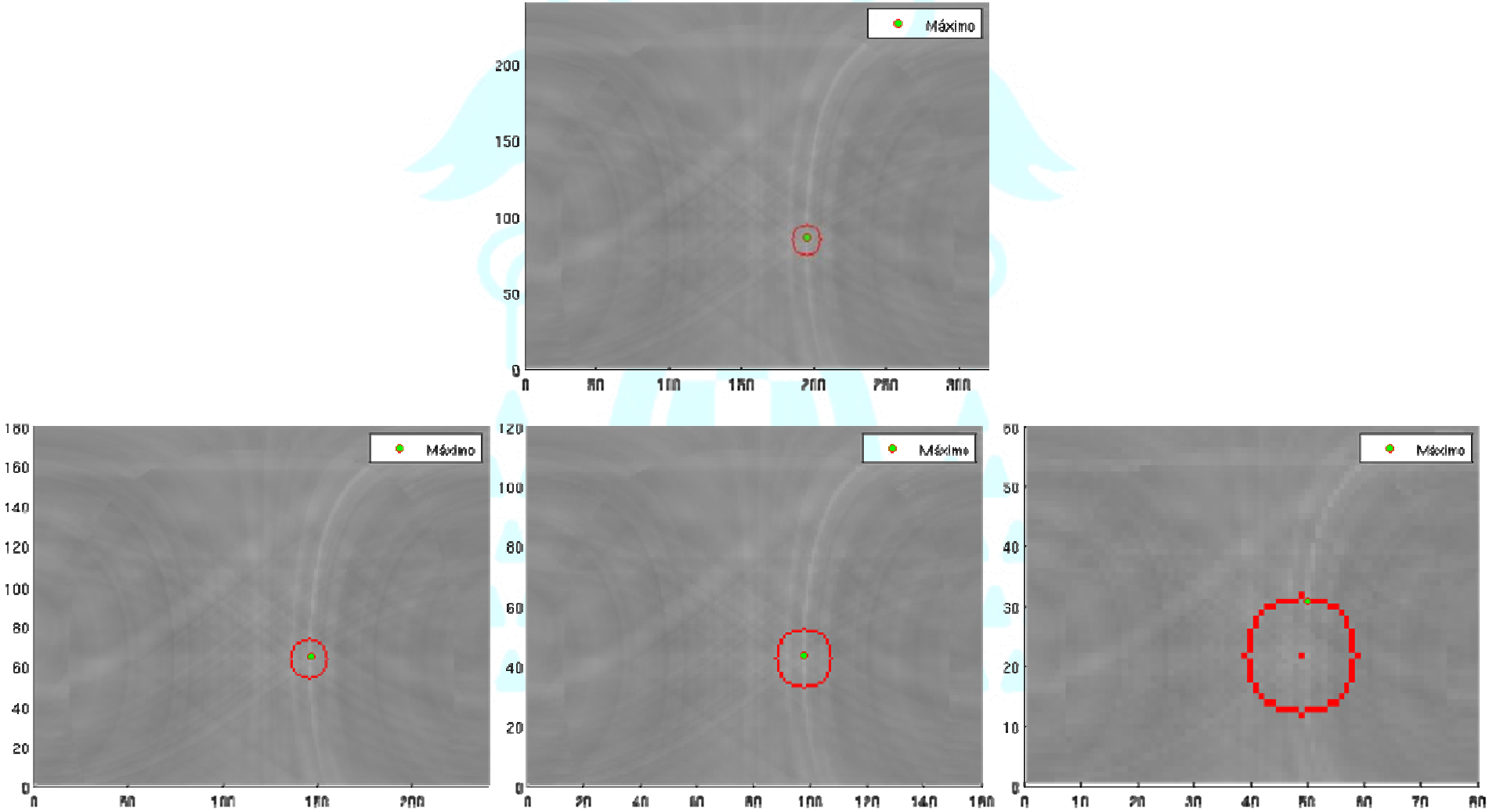
- Experimento base:
  - Error elevado en estimación DLT



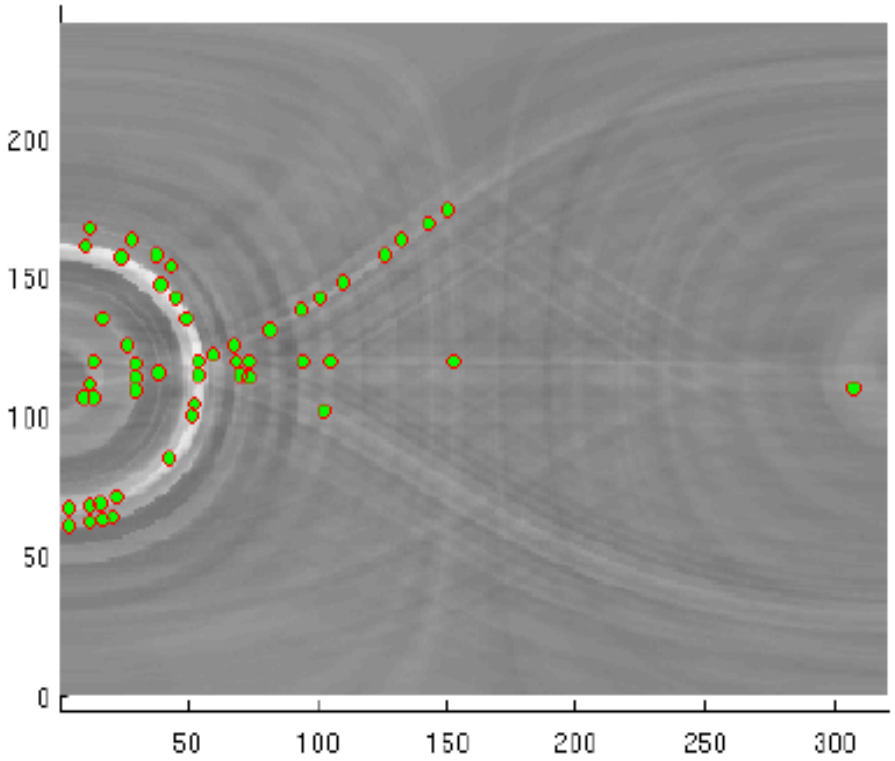
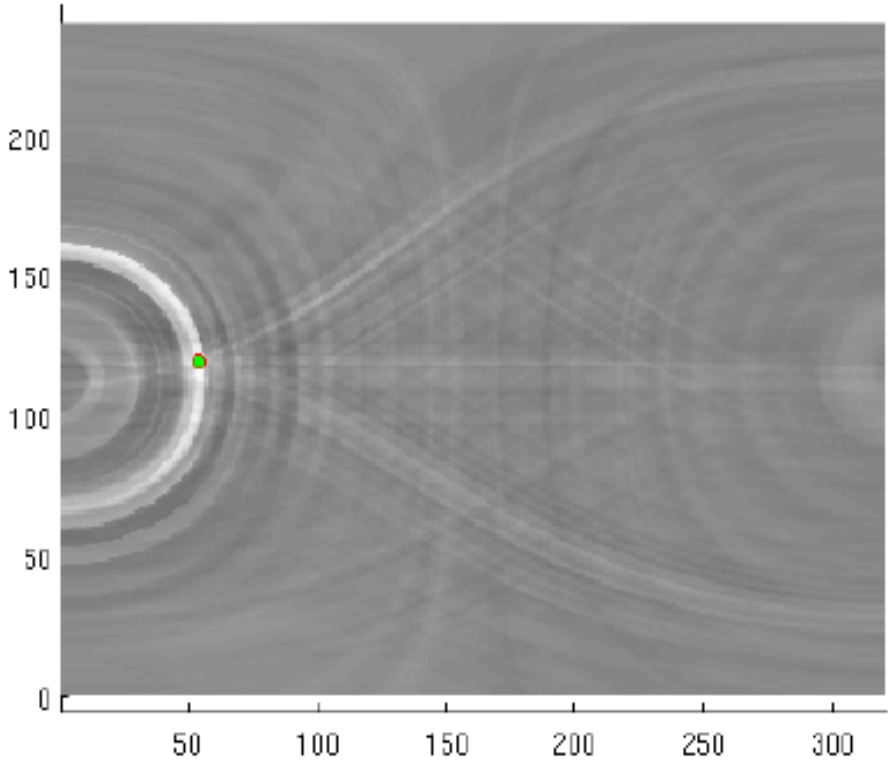
Evaluación experimental

Resultados

- Barrido en resolución:



Evaluación experimental

Resultados

- Non maximum supression:
  - Máximos locales
  - Problemas en selección de umbral