Memoria de Prácticas

Urban Multi-UAV Surveillance: A Dynamic Team Orienteering Approach

José Aguilar Camps
Grado en Ciencia de Datos, ETSINF
Universitat Politècnica de València

Tutores: Juan Miguel Alberola Oltra, Víctor Sánchez Anguix

1 Introducción

Esta memoria recoge el trabajo realizado durante las prácticas centradas en el diseño y evaluación de un sistema de coordinación autónoma de drones (UAVs) para vigilancia urbana persistente. El proyecto nació como un caso de uso concreto, la monitorización de puntos de interés turísticos en Altea, y evolucionó hacia un marco de optimización general, validado experimentalmente sobre redes urbanas reales de tres ciudades españolas de escalas crecientes.

El resultado final es el artículo *Urban Multi-UAV Surveillance: A Dynamic Team Orienteering Approach*, que introduce el *Team Orienteering Problem with Time-Dependent Rewards and Multiple Visits* (TOP-TD-MV), un marco que combina regeneración continua de la recompensa, visitas múltiples a un mismo punto de interés y restricciones realistas de energía y recarga.

2 Entidad receptora y contexto

Las prácticas se han desarrollado en el instituto VRAIN (Valencian Research Institute for Artificial Intelligence) de la UPV, en el marco de la Cátedra ENIA, bajo la supervisión de Juan Miguel Alberola Oltra y Víctor Sánchez Anguix, dentro de las líneas de investigación del instituto sobre optimización de sistemas multiagente y coordinación de flotas autónomas.

3 Objetivos

Los objetivos planteados al inicio de las prácticas fueron estudiar la coordinación de flotas de drones para vigilancia persistente, identificando las limitaciones de los enfoques clásicos de rutas frente a prioridades cambiantes en el tiempo, y diseñar un modelo formal con restricciones energéticas, infraestructura de recarga limitada y recompensa dinámica dependiente de la franja horaria y del tiempo desde la última visita. A partir de ahí, se buscó implementar un entorno de simulación en Julia capaz de generar instancias reproducibles a partir de redes de calles reales, diseñar y evaluar un planificador de referencia, tanto descentralizado como centralizado, mediante distintas métricas de puntuación, y finalmente redactar los resultados en forma de artículo científico.

Financiado por la Unión Europea NextGenerationEU

4 Planteamiento inicial del problema

La primera fase consistió en definir un caso de uso acotado: coordinar una flota de drones para monitorizar puntos de interés turísticos en Altea bajo restricciones energéticas. A diferencia de los escenarios clásicos de rutas, donde el objetivo es alcanzar un destino, aquí interesa maximizar la información obtenida a lo largo de un horizonte temporal, lo que exige optimizar tanto las rutas espaciales como la sincronización temporal de las visitas frente a prioridades dinámicas.

El escenario se abstrajo en tres agentes: el dron, que se desplaza por el grafo urbano recolectando recompensa bajo autonomía limitada y no puede iniciar un trayecto sin garantizar después una recarga segura; el punto de interés, cuya recompensa depende de la franja horaria y del tiempo transcurrido desde la última visita, reiniciándose al ser visitado; y el cargador, que no aporta recompensa pero permite la regeneración energética.

Como primera aproximación se implementó un sistema descentralizado en Julia en el que cada dron aplica una política *greedy* local: prioriza la carga si su batería es crítica, y si no, el punto de interés con mejor relación recompensa/distancia, reservando en exclusiva el recurso elegido. Este enfoque, robusto y de bajo coste, sirvió como referencia de partida, aunque al carecer de coordinación global tiende a soluciones subóptimas.

5 Formalización: TOP-TD-MV

La revisión de la literatura sobre el Team Orienteering Problem (TOP) reveló dos limitaciones para modelar vigilancia persistente: la recompensa solo se recoge en la primera visita, y se asume estática en el tiempo. Esto motivó formalizar el TOP-TD-MV.

El entorno se modela como un grafo no dirigido completo $G = (V, E)$, con tiempos de viaje t_{uv} en cada arista. Un subconjunto $P \subseteq V$ de N nodos son los POIs a vigilar, y la misión la lleva a cabo una flota homogénea U de M UAVs ($M \ll N$), sin obligación de visitar todos los POIs ni restricción a una única visita, dentro de un horizonte $T_{\max} = 24$ horas.

Cada POI p_i tiene una recompensa dinámica $S_i(t) = R_i(t) \cdot \sigma_i(\tau_i(t))$, producto de un potencial máximo dependiente del tiempo $R_i(t)$ (interpolación lineal de un perfil horario) y una recuperación dependiente de la visita $\sigma_i(\tau_i(t)) = \min(1, \tau_i(t)/T_{\text{cool}})$, que crece linealmente desde la última visita hasta el umbral de enfriamiento T_{cool}.

Cada UAV u_k tiene autonomía restante $b_k(t) \in [0, B_{\max}]$. Las estaciones de carga $C = V \setminus P$ tienen capacidad única y cola FIFO; un movimiento hacia un POI solo es factible si el dron conserva autonomía suficiente para alcanzar después alguna estación de carga. El objetivo es maximizar la recompensa total acumulada por la flota a lo largo del horizonte de misión, asumiendo observabilidad completa de recompensas y asignaciones.

6 Metodología de resolución

El planificador centralizado asigna, en cada paso de replanificación, cada UAV libre al POI que maximiza una métrica de puntuación M sujeta a factibilidad energética; si ninguno es factible, el dron va a la estación de carga más cercana (Algoritmo 1).

Se evaluaron tres métricas: Random (puntuación aleatoria, suelo de rendimiento), Reward ($M = S_p(t+t_{up})$, ignora el coste de desplazamiento) y Reward Ratio ($M = S_p(t+t_{up})/t_{up}$, que equilibra valor y coste de viaje).

Algorithm 1 Lógica de replanificación *greedy*

Require: UAVs libres U_{free}, POIs libres P_{free}, cargadores C

for cada $u \in U_{\text{free}}$ **do**
  $w^* \leftarrow -\infty$, $p^* \leftarrow \emptyset$
  for cada $p \in P_{\text{free}}$ factible energéticamente **do**
    $w_{up} \leftarrow M(u, p, t)$; si $w_{up} > w^*$: $w^* \leftarrow w_{up}$, $p^* \leftarrow p$
  end for
  Despachar u hacia p^* si existe, si no hacia el cargador más cercano
end for

7 Diseño experimental

Al no existir instancias estandarizadas, se diseñó un generador propio de escenarios reproducibles a partir de redes de calles reales de tres ciudades españolas (OpenStreetMap, componente conexa mayor), con matrices de tiempo de viaje precalculadas mediante Dijkstra a 50 km/h.

Cuadro 1: Entornos urbanos operativos

Ciudad	POIs ($\|P\|$)	Cargadores ($\|C\|$)	UAVs ($\|U\|$)
Elche	50	3	5
Valencia	100	5	10
Madrid	200	10	20

La flota es homogénea, con $B_{\max} = 5400$ s (1,5 h) y recarga completa en 1800 s (30 min, $\alpha = 0,5$). Cada POI recibió aleatoriamente uno de siete perfiles horarios de recompensa, con $T_{\text{cool}} = 3600$ s (1 h) global. Cada ciudad se instanció con 10 semillas, dando 30 instancias en total. Todo el marco se implementó en Julia 1.11 sobre un Intel Xeon Silver 4210 (32 GB RAM).

8 Resultados y discusión

Random define el suelo de rendimiento al no incorporar prioridad espacial ni temporal. Reward mejora este resultado, pero al ignorar el coste de desplazamiento agota la autonomía en pocas misiones costosas: menos de 1000 viajes en Elche y solo 1817,6 en Madrid, capando su recompensa en 8966,0. Reward Ratio resulta claramente ganadora en todas las escalas: en Madrid completa 8841,9 viajes (casi cinco veces más que Reward) y alcanza una recompensa media de 21766,2, unas 2,4 veces la de Reward. Al mantener trayectos cortos, dispersa la flota y sostiene una mayor proporción de POIs recientemente visitados, justo lo que exige la vigilancia persistente.

Estos resultados confirman que el compromiso explícito entre valor de vigilancia y coste de desplazamiento es el factor decisivo del rendimiento, y que la propia estructura de recompensa dinámica basta para guiar un comportamiento eficaz incluso bajo una política de decisión mínima.

9 Conclusiones

Partiendo de un caso de uso acotado (Altea), el trabajo evolucionó hacia el TOP-TD-MV, un marco que modela regeneración continua de la prioridad de monitorización, cargadores de capacidad única con cola FIFO y necesidad de visitas repetidas, validado sobre 30 instancias de tres

Financiado por la Unión Europea NextGenerationEU

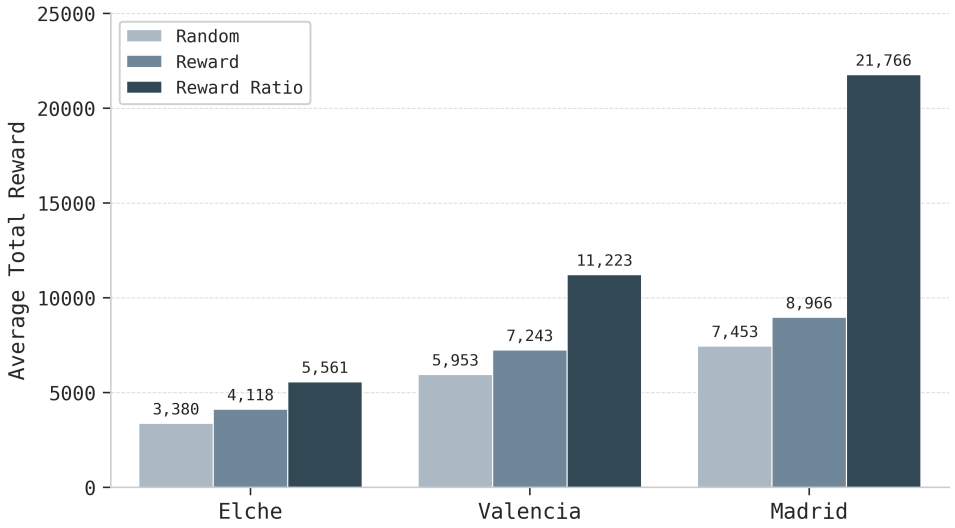

Figura 1: Recompensa media total por métrica y ciudad

ciudades reales. La métrica Reward Ratio ofrece de forma consistente el mejor rendimiento en todas las escalas, sin necesidad de ningún mecanismo de coordinación explícita entre drones.

10 Competencias adquiridas

A lo largo de las prácticas se ha trabajado en la formalización matemática de problemas de optimización combinatoria con restricciones energéticas, en la programación en Julia para simulación multiagente, y en el diseño de bancos de pruebas reproducibles a partir de datos geoespaciales reales. Igualmente se ha practicado el diseño experimental comparativo y el análisis crítico de resultados, así como la redacción científica en inglés según las convenciones del área.

11 Líneas futuras

Como continuación natural de este trabajo, cabría explorar métricas con anticipación temporal de la recompensa esperada en el instante de llegada, en lugar de limitarse a una mirada puramente espacial. También resultaría de interés incorporar la congestión de los cargadores directamente en la decisión de asignación, mitigando los cuellos de botella de sincronización de batería en despliegues densos. Más allá del planificador *greedy*, la estructura de recompensa del TOP-TD-MV ofrece una señal de entrenamiento natural para el aprendizaje por refuerzo multiagente, capaz de descubrir políticas no miopes que el enfoque actual no puede representar. Finalmente, quedaría abierta la extensión a flotas heterogéneas y a escenarios con incertidumbre y detección de eventos en tiempo real.

Agradecimientos

Se agradece el apoyo de la Cátedra ENIA del instituto VRAIN, UPV, y la supervisión de Juan Miguel Alberola Oltra y Víctor Sánchez Anguix durante estas prácticas.