



Università degli Studi di Salerno

Dottorato di Ricerca in Informatica e Ingegneria dell'Informazione
Ciclo 32 – a.a 2018/2019

TESI DI DOTTORATO / PH.D. THESIS
(ABSTRACT)

Context-aware knowledge extraction for UV scene understanding

DANILO CAVALIERE

SUPERVISOR: **PROF. SABRINA SENATORE**

PHD PROGRAM DIRECTOR: **PROF. PASQUALE CHIACCHIO**

Dipartimento di Ingegneria dell'Informazione ed Elettrica
e Matematica Applicata
Dipartimento di Informatica

Abstract (English)

In the surveillance systems, Unmanned Vehicle (UV) scene interpretation is a non-trivial problem, because UVs need to possess human-like common-sense knowledge to correctly interpret events and situations occurring in the monitored environment. Mobile camera-related issues, such as motion blur, can further complicate scene interpretation, causing a lack of reference points that badly affects the interpretation of scene entities and situations. To this purpose, this thesis investigates the synergistic combination of video tracking with Semantic Web technologies to enhance UVs at the interpretation of dynamic scenarios.

The first part of the thesis provides a survey conducted on the methods employed to extract knowledge from the acquired structured and unstructured data. When dealing with unstructured data, there is the need to define and process contextual data to extract high-level concepts from text. To this purpose, an approach is introduced to mine concepts from texts by building layered contextual knowledge on document terms exploiting a geometrical structure, called Simplicial Complex. Then, the focus switches to the knowledge extraction from multimedia data, and more specifically, video data. To this purpose, an ontology-based approach is presented to represent the video scene as composed of mobile (i.e., people, vehicles) and fixed entities (i.e., environmental sites and features), along with the spatio/temporal relations among them. The use of the ontology reasoning can support alerting event detection.

In the second part, solutions to detect high-level activities and events have been investigated. In order to build high-level descriptions of the activities and events, there is the need to build knowledge on various aspects of the scene at various levels of detail (scene object, environment, specific events, overall situation), additionally, there is a lack of general and reusable models to build

knowledge on the various levels. In literature there are lots of approaches that fuse data to spot simple or contextual events, but there are no models to relate them in order to fully describe the scene. Therefore, a composing approach is proposed to recognize complex activities from simpler activities carried out by the mobile entities in the scene. Then, a comprehensive situation detection approach is introduced. It defines a multi-ontology design pattern to incrementally build various layers of knowledge and depict the whole scene observed by the UVs, from the single scene entities to the high-level activities and situations.

The last part of the thesis analyses the scenario interpretation through systems employing multiple UVs and smart devices. Multiple UVs need ways to combine the knowledge they acquired to better interpret what happened. To this purpose, the previously introduced ontology-based framework has been enhanced with a novel agent-based model to allow UVs to cooperatively build knowledge on the scene. The employment of multiple devices can indeed provide better views on the scene to monitor, however, UVs not naturally come to an agreed scene interpretation. To tackle this issue, a consensus-based Group Decision Making (GDM) approach is proposed to support teams of UVs to robustly interpret the monitored scenario and evaluate the reliability of their interpretations.

Abstract (Italiano)

I veicoli autonomi (Unmanned Vehicles o UVs) vengono sempre più usati per monitorare un ambiente nell'ambito dei sistemi di sorveglianza. Per il monitoraggio automatico, gli UV necessitano di comprendere cosa accade nella scena. Tale compito non è banale in quanto richiede macchine in grado di ragionare come essere umani per interpretare eventi e situazioni. Inoltre, siccome tali veicoli impiegano telecamere per il monitoraggio della scena, sono soggetti a problemi legati al video, come lo sfocamento (motion blur), che mina un corretto rilevamento degli oggetti della scena e fa perdere punti di riferimento complicando l'interpretazione di eventi e situazioni. Di conseguenza, l'obiettivo di questa tesi è analizzare l'impiego sinergistico della Computer Vision con le tecnologie del Web Semantico per astrarre conoscenza dai dati video e supportare l'interpretazione di scenari dinamici.

La prima parte della tesi fornisce una panoramica dei metodi per estrarre conoscenza da documenti strutturati (structured data) e non strutturati (unstructured data). Quindi, vengono analizzati metodi per l'estrazione di conoscenza, in termini di concetti di alto livello, da testo in linguaggio naturale. L'estrazione di concetti richiede modelli per definire e processare dati contestuali. A tal proposito, viene introdotto un approccio che definisce lo spazio topologico linguistico dei termini estratti dal testo attraverso una struttura geometrica chiamata Simplicial Complex, la quale consente di generare conoscenza contestuale e stratificata per supportare il rilevamento di concetti dal testo. Successivamente, vengono analizzate metodologie per estrarre conoscenza da dati multimediali, ed in particolare dati video. In tale ambito, viene proposto un modello ontologico della scena video che include oggetti mobili, quali veicoli, persone ecc., ed oggetti fissi, come siti ed elementi dell'ambiente (strade, parchi, bancomat, ecc.). Il modello rappresenta anche le relazioni spazio/temporali tra i vari

oggetti, e attraverso l'inferenza, può generare ulteriore conoscenza a supporto del rilevamento di eventi.

Nella seconda parte della tesi, vengono investigate soluzioni per il rilevamento di eventi e attività da una scena video. La comprensione di eventi e attività richiede modelli in grado di generare conoscenza a diversi livelli della scena (livello oggetti, livello ambiente, eventi specifici e situazioni complessive). Inoltre, vi è la necessità di modelli riusabili e generali per generare conoscenza a più livelli. In letteratura, vengono proposti diversi approcci che sfruttano dati contestuali per rilevare eventi semplici, ma nessuno di essi astrae da tali eventi per sintetizzare descrizioni ad alto livello della scena. A tal proposito, viene introdotto un modello di composizione, che integra eventi e attività semplici effettuate dagli oggetti mobili della scena per ottenere attività complesse. In seguito, viene introdotto un modello di comprensione della scena complessiva. Tale modello consiste in un design pattern ontologico che sfrutta diverse ontologie di alto livello per generare conoscenza su ogni livello della scena, dai singoli oggetti della scena alle attività e situazioni, astruendo conoscenza per ottenere una descrizione sintetica della scena.

L'ultima parte della tesi approfondisce lo studio di modelli di interpretazione della scena che impiegano dispositivi multipli, includenti UV e sensori fissi. Dispositivi multipli necessitano di modelli per acquisire e integrare la conoscenza acquisita da ciascun di essi per interpretare in modo efficiente ed univoco la scena. A tale scopo, il modello ontologico, introdotto in precedenza, viene riusato per ciascun dispositivo ed un modello ad agenti viene introdotto per consentire ai dispositivi di generare conoscenza in modo cooperativo sulla scena. L'impiego di più dispositivi fornisce una visione plurale e più completa di tutta la scena, ciononostante, in base ai dati acquisiti e processati ciascun dispositivo nel team può avere una sua interpretazione della scena osservata. Di conseguenza, i dispositivi non ottengono un'interpretazione comune della scena. Per affrontare tale problema, è stato proposto un modello basato su metodologie di Group Decision Making (GDM) per consentire ai dispositivi di ottenere un'interpretazione comune della scena,

ed inoltre stimare quanto tale valutazione sia attendibile.