

Natural Vision Systems have reached incredible performances in detecting and tracking multiple moving objects simultaneously. Accurate and robust multiple-target tracking is also a key task in many promising Computer-Vision applications. Practical usages of proposed algorithms can now be tackled in real time thanks to recent technological advances. Further, this represents a huge challenge because of the numerous particular problems involved in such a task. Thus, proposals must deal with multiple highly non-rigid targets which move in an unforeseeable manner through unconstrained dynamic open-world scenarios.

In this thesis, a principled hierarchical architecture which fulfills multiple-target tracking is presented. Further, another tracking approach—based on *particle filtering*—is previously developed and evaluated. Thus, a modular and hierarchically-organised system is designed. It is conformed by a detection level which feeds a two-level tracking subsystem. Co-operating modules, distributed through this architecture, work following both bottom-up and top-down approaches. Contributions include both the architecture itself, and the development, improvement and integration of the different modules. The proposed architecture introduces the necessary synergies which allow the system to tackle such a problem as unconstrained multiple-target tracking. With respect to the different modules, the main focus is placed on high-level tracking algorithms. Since a careful analysis of motion events is a critical issue for tracking successful, a module for principled event management is proposed, and embedded in the system. Multiple-target interaction events, and a proper scheme for tracker instantiation and removal according to scene events, are considered. Thus, the system is allowed to switch among the two different operation modes implemented, motion-based tracking and appearance-based tracking. This entails another remarkable characteristic of the system: its ability to continuous and independently track numerous targets while they group and split. Multiple appearance models are built and constantly updated. A special attention is paid to maximise the discrimination between the target and potential distracters by means of an appropriate feature selection, and a wise combination of all available sources of information.

It works as a stand-alone application in a non-friendly, complex and dynamic scenario. No a-priori knowledge about either the scene or the targets, based on a previous off-line training period is needed. No camera calibration is required since tracking is achieved without the need of 3D information.

Successful tracking has been demonstrated in multiple sequences of both indoor and outdoor scenarios. Accurate and robust localisations have been yielded even during long-term target clustering and occlusions. Results are comprehensively analysed.

Los Sistemas de Visió Natural (SVN) han assolit uns resultats increïbles pel que fa a la detecció i seguiment de múltiples objectes simultàniament en moviment. Aquest seguiment precís i robust de múltiples agents (objectes i persones) és també una tasca clau en moltes aplicacions prometedores basades en la Visió per Computador (VC). Los algorismes teòrics proposats durant estos últims anys poden ser ara aplicats a la pràctica i en temps real gràcies als últims avenços tecnològics. Nogensmenys, això ha representat ser un gran repte degut als nombrosos problemes que han anat sorgint durant el desenvolupament d'aquesta tasca, bàsicament pel fet d'haver de tractar amb múltiples persones, que són altament no-rígid, que es mouen d'una manera imprevisible per escenaris oberts, dinàmics i no-restringits.

En esta Tesis se presenta una arquitectura jeràrquica que realitza el seguiment de múltiples agents. Les contribucions inclouen l'arquitectura en si i el desenvolupament, millora i integració dels diferents mòduls. L'arquitectura proposada introdueix les sinergies necessàries per permetre al sistema tractar el problema del seguiment de múltiples agents. Respecte als diferents mòduls, el focus principal es posa en els algorismes de seguiment d'alt nivell. Ja que una anàlisi prudent dels esdeveniments de moviment és un assumpte crític per un seguiment d'èxit, es proposa un mòdul per a la gestió d'aquests esdeveniments, que està incrustat en el sistema. Així, es consideren els esdeveniments d'interacció entre múltiples agents, i un esquema propi per a la instanciació dels diferents algorismes de seguiment o la seva supressió segons els esdeveniments de l'escena. Així, lo sistema permet canviar-se entre dos modes diferents de funcionament implementats, és a dir, basat en lo moviment i basat en l'aparença. Això suposa una altra característica notable del sistema: la seva habilitat per seguir continua i independentment múltiples objectius mentre aquests s'agrupen i es separen. Es construeixen models d'aparença que s'actualitzen constantment. Es para una atenció especial per maximitzar la discriminació entre l'objectiu seguit i los distractors potencials, per mitjà d'una selecció de les característiques més apropiades i una combinació assenyada de totes les fonts d'informació disponibles.

Aquesta arquitectura de seguiment treballa com a aplicació autònoma en un escenari no amistós, complex i dinàmic. No es necessita cap tipus de coneixement a-priori al voltant de l'escena o dels agents, basat en un període d'entrenament previ. Per això, l'escenari podria ser completament desconegut per endavant. No s'exigeix cap calibració de càmeres ja que lo seguiment és aconseguit sense la necessitat d'informació 3D.

S'ha demostrat un seguiment correcte en múltiples seqüències de tant interiors com a l'aire lliure, de bases de dades pròpies i públiques molt conegudes. Han estat assolides unes localitzacions acurades i robustes fins i tot durant agrupaments llargs i oclusions. Los resultats obtinguts s'han analitzat extensa i completament.