Expert Systems With Applications 87 (2017) 118–128

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

# Agent-based framework to individual tracking in unconstrained environments

Cauê Zaghetto, Luiz Henrique M. Aguiar, Alexandre Zaghetto, Célia G. Ralha, Flavio de Barros Vidal*

*Biometrics and Technologies Group, Department of Computer Science, University of Brasilia, Campus Universitario Darcy Ribeiro, Asa Norte, Brasilia, DF, Brazil*

ARTICLE INFO

*Article history:*
Received 29 August 2016
Revised 26 May 2017
Accepted 27 May 2017

*Keywords:*
Multiagent systems
Unconstrained tracking
Cloud storage

ABSTRACT

Agents with intelligent perceptual capabilities are considered state-of-the-art in advanced intelligent systems. In addition, a multiagent system is considered an enabling technology for applications that rely on distributed and parallel processing, including data, information and knowledge in complex computing environments. With the aim of creating advanced intelligent systems with visual perception, this paper presents an agent-based framework to individual tracking in unconstrained environments. The framework has three types of agents that interact using the Contract Net Protocol. The face detector and tracker agents perform fully automatic single-sample face recognition using the Viola-Jones and the Scale Invariant Feature Transform/Speeded Up Robust Features algorithms. The experimental results show that the framework adequately recognizes and tracks individuals in unconstrained environments, indicating the path the individuals have taken and the time they spent in the field of view of the surveillance agents. Some of the open source framework advantages are the distribution in heterogeneous infrastructures, the expansion with new agents using different face recognition algorithms (e.g., Eigenfaces), and the individual tracking logs that can be used in different ways, e.g., improve security in surveillance areas such as automated teller machines, self-paying kiosks, movie box offices, and malls.



## 1. Introduction

Artificial Intelligence (AI) research has evolved from different paradigms since its birth at the Dartmouth Conference in 1956 (Luger, 2008). From the classical to the modern approach, where computational intelligence is the study of the design of intelligent agents (Poole, Mackworth, & Goebel, 1997), AI is concerned with intelligent behavior in artifacts (Nilsson, 1998). Thus, modern AI deals with the design of intelligent systems with rational agents that can perceive and act upon the environment (Russell & Norvig, 2010). From this modern paradigm, new perspectives of AI study has emerged where traditional concepts such as pattern recognition, data mining, and numerical optimization are types of algorithms grounded to an environment as a systems approach using the sensor–algorithm–effecter view (Jones, 2008).

For the next-generation AI, known as the AI 2.0, the advanced intelligent systems have to be empowered with intelligent perceptual capabilities. Perception is cited as the interaction interface between intelligent systems and the real world. Perception is thus the most significant capability to empower intelligent systems. The state-of-the-art across different areas of perception in the AI 2.0 era includes visual perception (Pan, 2016; 2017; Tian et al., 2017).

Considering the research and development trends in visual perception of AI 2.0, this work investigates how to define and develop a robust agent-based framework with face detection and identification that can manage individual tracking in real time to be applied in unconstrained environments. Our hypothesis is that without a flexible visual perception feature it is impossible to create advanced intelligent systems. Perception in an intelligent system begins with sensor data in various forms that is processed along with prior knowledge and models to extract relevant information to the task of agents in the AI system. Therefore, data from perception forms situational awareness that provides agents with a comprehensive knowledge about the state of the world necessary to understand, plan and execute tasks effectively and safely.

As a transformative perception technology, this work focuses on the automatic verification or recognition of an individual, based on physiological and behavioral characteristics through biometric systems (Jain, Bolle, & Pankanti, 1998; Wayman, Jain, Maltoni, & Maio,

* Corresponding author.
*E-mail addresses:* zaghettoc@bitgroup.co (C. Zaghetto), luiz@bitgroup.co (L.H.M. Aguiar), zaghetto@unb.br (A. Zaghetto), ghedini@unb.br (C.G. Ralha), fbvidal@unb.br, fbvidal@cic.unb.br (F.d.B. Vidal).

http://dx.doi.org/10.1016/j.eswa.2017.05.065
0957-4174/

2010). The current state-of-the-art face biometric algorithms performance is encouraging, as far as identification in controlled conditions is concerned (Sarkar et al., 2005; Zhao, Chellappa, Phillips, & Rosenfeld, 2003). Thus, face recognition in constrained environments, especially considering the advances in deep learning, has achieved great success. Nevertheless, in more realistic scenarios, such as in unconstrained environments, it needs more research to effectively exhibit intelligent perception.

In this respect, the use of multiagent systems (MAS) is adequate. MAS are made of multiple interacting intelligent agents with some degree of autonomy, being able to cooperate, compete, communicate and act flexibly to achieve defined objectives. MAS are considered an enabling technology for applications that rely on distributed and parallel processing, including data, information and knowledge in complex computing environments (Weiss, 2013). In addition, the multiagent technology cuts across several widely-recognized computing paradigms, such as grid computing, autonomic computing, service-oriented computing, cloud computing, involving open, large-scaled and networked resources.

With the aim of creating advanced intelligent systems with visual perception, this paper presents a novel agent-based framework to extract relevant information to proceed towards the task of detecting/identifying faces and individual tracking in unconstrained environments. The framework has three different types of agents: manager, face detection, and tracking that interact using the Contract Net Protocol. The face detection agent performs a single-sample face recognition that is important to many real-world applications, such as federal agencies, commercial organizations, and banks, as there is only a single facial image per subject available. The face detection and identification process is fully automatic using the Viola–Jones (Viola & Jones, 2004) and the Scale Invariant Feature Transform/Speeded Up Robust Features (SIFT/SURF) algorithms (Lowe, 1999; 2004). The tracking agent uses information registered by the face detection agent.

As the main contributions of this work we may cite:

- The definition and implementation of an open source framework with visual perception capabilities to build upon the creation of advanced intelligent systems;
- An agent-based framework that adequately recognizes and tracks individuals in unconstrained environments, indicating the path each individual has taken and how much time he spent in the field of view of the surveillance agents;
- The framework is essentially distributed in heterogeneous computational infrastructures equipped with one camera per agent able to cover broader environments;
- The framework is expandable to include new agents, allowing the use of different face recognition algorithms (e.g., Eigenfaces) in the same architecture;
- The agent communication process is done using the Foundation for Intelligent Physical Agents (FIPA)-Agent Communication Language (ACL) through a shared directory in the cloud;
- The agent interaction follows a negotiation standard using the Contract Net Protocol.

The remainder of this paper is organized as follows: Section 2 presents the most related conceptual background, including intelligent agents and MAS, face detection and face identification; in Section 3 related works are presented; in Section 4 an overview of the proposed agent-based framework is presented, including the agents design, architecture, communication and interaction protocols, and the framework workflow; in Section 5 the experimental results are presented; and in Section 6 the conclusions and future work are discussed.

## 2. Background

In this section, an overview of concepts related to intelligent agents and MAS, face detection and face identification will be presented.

### 2.1. Intelligent agents and MAS

According to Wooldridge (2009) and Weiss (2013), MAS are systems composed of multiple interacting computing elements called agents. An agent is a computational entity situated in some environment being able to perceive its states and act upon it. As intelligent entities, agents have autonomous capabilities to act with flexibility, according to a variety of environmental circumstances. Agents can use diverse processes, such as searching mechanisms, constraint satisfaction, planning and learning skills. Thus, intelligent agents are to some extent able to act autonomously in order to achieve their design objectives and interact with other agents, not simply exchanging data but actively engaging in cooperative and/or competitive scenarios. In summary, intelligent agents are capable of social interactions, analogous to humans daily activities, involving communication (e.g., semantically rich languages) and making decisions based on norms, negotiation, argumentation, voting, auctioning, and coalition formation.

In Wooldridge (2009) there is a formal definition of the abstract concept of an agent. We can assume that each agent has a finite set of actions $Ac$ to transform the states of the environment $Ac = \{\alpha_0, \alpha_1, \alpha_2, \ldots\}$. The environment $E$ is formed by a discrete set of states $E = \{\epsilon_0, \epsilon_1, \epsilon_2, \ldots\}$. The run $r$ of an agent in an environment is a sequence of states and actions $r : \epsilon_0 \xrightarrow{\alpha_0} \epsilon_1 \xrightarrow{\alpha_1} \ldots \xrightarrow{\alpha_{n-1}} \epsilon_n$, while $R$ is the set of all possible finite sequences over $E$ and $Ac$. $R^{Ac}$ and $R^E$ are subsets of $R$ that end with an action or an environment state, respectively. The effect that an agent's actions have on the environment is represented by the state transformer function $\Gamma$: $R^{Ac} \rightarrow \wp(E)$. This function maps a run that ends with an action to a set of possible environmental states. In addition, the environment is assumed to be history-dependent, which means that earlier actions also play a part in determining the current state in the environment. Since agents are autonomous, there is uncertainty about the result of performing an action in some state (nondeterminism). If $\Gamma(r) = \emptyset$ and $r$ ends with an action, the system has ended its run. Thus, an environment can be represented by a triple $Env = < E, \epsilon_0, \Gamma >$. Based on the presented definitions, we say that two agents $Ag_a$ and $Ag_b$ are behaviorally equivalent with respect to $Env$, if $R(Ag_a, Env) = R(Ag_b, Env)$.

In order to exchange and understand messages, agents in MAS must use an ACL. In literature, there are many proposed agent languages, e.g., speech acts (Searle, Kiefer, & Bierwisch, 1980), knowledge query and manipulation language (KQML)/knowledge interchange format (KIF), and the ACL standard of FIPA (2004). In general, an ACL can be seen to have two main components: a performative verb (request, inform, inquire) and a propositional content (propose, accept, reject, retract, disagree, counter-propose a course of action). In FIPA-ACL there are many performatives with a basic structure including information of sender, receiver, content, language, and ontology (i.e., specific concepts and relationships of a domain). The performatives are used to pass or request information, negotiate, perform actions, or error handling (failure, not-understood). The communication protocol of MAS can be specified at several levels: lowest – specifies the method of interconnection; middle – specifies the format (syntax) of information being transferred; and top – specifies the meaning (semantics) of the information.

Another well-known important requirement of MAS is the interaction protocol (Russell & Norvig, 2010; Weiss, 2013; Wooldridge, 2009). Agent interaction protocols enable agents to