

A vision and speech enabled, customizable, virtual assistant for smart environments

G. Iannizzotto, L. Lo Bello, A. Nucita, G. M. Grasso

Recent developments in smart assistants and smart home automation are lately attracting the interest and curiosity of consumers and researchers. Speech enabled virtual assistants (often named smart speakers) offer a wide variety of network-oriented services and, in some cases, can connect to smart environments, thus enhancing them with new and effective user interfaces.

However, such devices also reveal new needs and some weaknesses. In particular, they represent faceless and blind assistants, unable to show a face, and therefore an emotion, and unable to 'see' the user. As a consequence, the interaction is impaired and, in some cases, ineffective. Moreover, most of those devices heavily rely on cloud-based services, thus transmitting potentially sensitive data to remote servers.

To overcome such issues, in this paper we combine some of the most advanced techniques in computer vision, deep learning, speech generation and recognition, and artificial intelligence, into a virtual assistant architecture for smart home automation systems.

The proposed assistant is effective and resource-efficient, interactive and customizable, and the realized prototype runs on a low-cost, small-sized, Raspberry PI 3 device.

For testing purposes, the system was integrated with an open source home automation environment and ran for several days, while people were encouraged to interact with it, and proved to be accurate, reliable and appealing.