Robust Face Detection and Japanese Sign Language Hand Posture Recognition for Human-Computer Interaction in an "Intelligent" Room

†Jean-Christophe Terrillon, †Arnaud Pilpré,
†Yoshinori Niwa and ‡Kazuhiko Yamamoto
†Office of Regional Intensive Research Project (HOIP), Softopia Japan Foundation,
4-1-7 Kagano, Ogaki-City, Gifu 503-8569, Japan
{terrillon, pilpre, niwa}@softopia.pref.gifu.jp
‡Faculty of Engineering, Gifu University,
1-1 Yanagido, Gifu-City, Gifu 501-1193, Japan
yamamoto@info.gifu-u.ac.jp

Abstract

A system for the detection of human faces and for the classification of hand postures of the Japanese Sign Language in color images inside an "intelligent" room is presented. We first propose to apply a combination of a skin chrominance-based image segmentation with a color vector gradient-based edge detection [1] [2] to efficiently detect faces and hands. Within the framework of a general approach, a statistical model for face detection based on invariant moments [3] [4] is used to discriminate between faces and hands in the segmented images. A novel approach to hand posture recognition based on phase-only correlation [5] is then adopted to classify a subset of static hand postures of the Japanese Sign Language, each posture representing a given phoneme, and also to discriminate between hand postures and the image scene background. Experiments show that the additional use of the color vector gradient significantly improves the correct rate of face detection, and that the phase-only correlation filter yields a high rate of discrimination between different static hand postures as well as between hand postures and the scene background. Ultimately, the system is to contribute to the implementation of meaningful human-machine interactions in a room that we are in the process of establishing, the "percept-room", mainly for welfare applications.

1. Introduction

Recently, so-called "intelligent" environments, in which a range of human activities can be automatically sensed, analyzed and "understood" by use of various computer vision technologies that are the least conspicuously embedded in the environment but are ubiquitous, have been developed [6] [7] [8] [9]. Meaningful human-machine interactions in an "intelligent" room such as the "percept-room", which we are in the process of implementing by use of multiple cameras [10], require as a first step the automatic detection of human faces, as well as of hands, for higher-level face and gesture recognition tasks. In particular, such human-machine interfaces allow a human user to control a variety of devices without any physical contact with remote controls, keyboards, etc... Various applications have been suggested, such as the contact-less control of home appliances (for example, of a television set, as in [11] [12]) for welfare improvement.

A fundamental issue to address is the level of complexity of the scene background that is to be expected in an "intelligent" room, because the robustness of the simultaneous detection and discrimination of faces and of hands (or recognition of hand postures) against complex scene backgrounds is a difficult problem which, to our knowledge, has not yet received much attention. Much work has focused on the robust detection of faces only (for example, [3] [4]), or of hands only [13], or on the robust recognition of (static) hand postures only [14] [15]. Also, it is often implicitly assumed that a face or a hand is present in a scene image. Finally, the "background" may also be considered to include the clothes that a person is wearing, other body parts than faces or hands (such as the neck and arms), and facial attributes such as glasses, hair and hairstyle, etc...

In this paper, we propose a system for the robust simultaneous detection of human faces and classification of hand postures of the Japanese Sign Language (JSL) in color images. The system can adapt to varying degrees of scene background complexity in indoor environments (office, home), to slowly varying illumination conditions, and it does not imply any a priori assumption about the presence of a face (or of more than one face) or of a hand (posture) simultaneously in an image. The system first uses a statistical skin color model to segment images and a statistical regularity-based shape model to detect faces. We then apply, to our knowledge for the first time, phase-only