Automatic detection of high risk situations for elder persons care

José M. Cañas, Antonio Pineda, Pablo Barrera
Departamento de Ingeniería Telemática y Tecnología Electrónica
Universidad Rey Juan Carlos
28933 Móstoles
jmplaza@gsyc.escet.urjc.es

Abstract

Artificial vision provides an incomparable good sensor when developing video care applications. They are not too intrusive and also supply a great amount of information, besides they are quite cheap. A video care application must work in real time providing tools for agile management of images. Furthermore, images only give us two-dimensional data but the most part of useful information is three-dimensional, so we must be able to extract it from them. In this article we introduce an application for elder persons care using only visual information. The system allows detecting problems such as a faint, a fall to the ground or reaching a window. The automatic monitoring of these people can provide them with a greater level of autonomy.

1. Introduction

The care of ill or elder persons implies a continuous watching of their daily tasks. In many cases their own families or the social services are in charge of their care in their own home or in specialized residences. But even counting with the necessary amount of people for their care, it is impossible to watch these patients continuously in order to detect as fast as possible any sort of incidents. The problem worsens for people who live alone, as they would need much more this type of assistance.

Video care traditional systems use terminals or special devices carried by the patients, with the shape of a bracelet or a necklace, which they activate when they need assistance, to send an emergency call to the appropriate services. Nevertheless, these systems require human interaction to report an alarm or ask for help. In the burden situations, it won't be possible for the patient to activate the device, with the consequence of a high-risk situation. On the other hand, the fact of having to carry this device continuously becomes a nuisance for the user because of the intrusive nature of the mechanism.

An alternative for this kind of video care service is using external monitoring devices. A set of video cameras can watch the patient automatically. When it makes a signal or, more important, when it watches some anomalous behavior such as a falling to the ground, the system automatically can send an emergency signal.

In the last years, this kind of automatically monitoring systems based on artificial vision has won importance. For example, in sanitary surroundings there are visual systems to detect anomalies in x-rays, and in biomedical environments there are 3D location systems to evaluate deviations in the column or to improve athlete's movements. Partly this is due to the increase of the computing ability of computers and the decrease of the price of video cameras, mainly because of their mass production. Nowadays, it is very common to find multitude of video cameras around us, they can be found in the surveillance systems, mobile phones, computers, videogame consoles and more.

Even with so many cameras available nowadays there are few applications totally independent based on artificial vision only. The main reason is that it is very hard to extract useful information from the cameras, which results in an increase of the price of the development. For the system to be useful in monitoring of ill persons it must be able to work as agile as possible with the purpose of turning on the alarm as fast as it can. Furthermore, it has to work 24x7. Reaching this goal with limited resources imposes some restrictions to the sort of system to develop and to the technology we can use.

Finally, we must consider the kind of information that becomes useful in a dangerous situation system detector as the one introduced here. In many cases one of the main problems when trying to identify dangerous situations consists of using visual information based in two dimensions. While analyzing images it can be very complex to find or decide between one situation or another. One example would be a person near or not to an ignited oven. If we only use a flat image there is ambiguity when estimating the distance so we could easily make a mistake. Avoiding most of these mistakes, when possible, results in a high computational cost of the application.

In this article we introduce an alternative that is focused on the extraction of the three-dimensional information in the scene. A great part of the useful information in our surrounding world is eminently three-dimensional, as it can be the relative position of an object opposed to another or a person's movement. When using 3D data it becomes simpler to describe the high-risk situations so the application is simpler and its only complexity is the extraction of three-dimensional information from the images.

Using this three-dimensional information and a set of simple rules defined over a 3D world, we have developed an application able to detect in real time high risk situations such as: a faint, a fall to the floor or reaching a window. Every part of it is implemented with a set of low cost cameras and a conventional PC.

The system we have developed has been designed to work with few computing resources and in heterogeneous environments. Although there are applications that only make use of the three-dimensional information from images [1], the classical methods to extract the information usually imply great computing costs. Partly for this reason, its use has only been kept in controlled environments or with great resources, as it can be industrial production lines or visual effects creation for movies or videogames.

2. Automatic detection of high risk situations

When an ill or an elder person finds herself alone it may be useful to count with a system that watches her with the purpose of detecting high risk situations as fast as possible. One system like this can help to raise her level of autonomy, as she doesn't depend on her family or qualified personal to permanently keep an eye on her.

The developed system is indeed oriented to elder persons that are susceptible of suffering falls or faints but don't want to live in a residence, as they prefer their own houses, enjoying the greater quality of life that it implies. A system like the one that is being introduced here allows a quick reaction when those high-risk situations (falls with loss of consciousness, faints, etc.) occur, situations that nowadays video care systems that use bracelets or necklace with a switch can't detect.

For an application like this to be really useful it must be fully autonomous. This means it must work without the necessity of a person supervising permanently the system, no matter if she is in the same place as the user or in one main and remote station. Our system achieves this since everything is automated. This makes families or watchers comfortable with the system, as they don't need to be permanently keeping an eye on the user. In fact, external care will only be necessary to attend the emergencies detected by the application, but not continuously.

In addition to this, the use of cameras in a house raises an important ethic dilemma related to privacy. With an automatic system as the one we have designed, privacy is kept, as nobody is watching the user, there is no one observing on "the other side". Images that are taken by the system are not stored anywhere and they remain inside the application except when a justified high-risk situation arises and the life of the person can be in great danger. This system can be described as an automatic alarm system, very similar to the volumetric sensors that are present in many houses and indeed detects our presence in different areas of the house.

Despite of this, as the cameras are great and powerful sensors, and there is a possibility of sending images remotely to external qualified personal, it will have to incorporate security mechanisms that ensure the protection against intrusions. If the alarm signal is reduced to a phone call or an SMS mobile message an invasion of the privacy won't be possible.

The basic system we have developed consists of several modules connected between them. We can see on Figure 1 its block diagram with the different modules that make up our application.

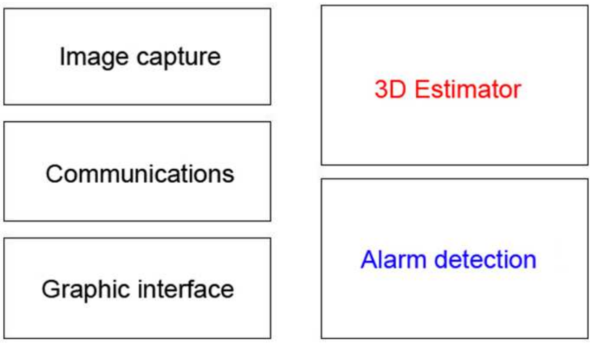

Figure 1: Block diagram of our elder care application.

First of all, we have an image capture system with a set of cameras, that can be static or mobile, that allows the continuous monitoring of the user. The position of the cameras for a typical application can be seen in Figure 2. In that figure the cameras are placed on every upper corner of the room so the system can see anything inside it. The cameras are connected directly, or through a private net, to a conventional desktop PC.

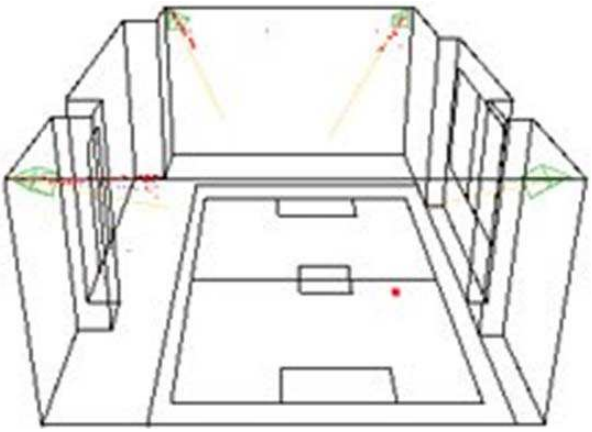

Figure 2: Setup of a prototype with 4 cameras at the room corners.

There is a module running in the PC, which is in charge of collecting all the images, processing them separately and merging the results all together to bring 3D information.

Once the person is already located in the scene we can use such information to decide whether that situation is dangerous or not. There are at least two main groups of risky situations. First, those related to dangerous places, like a balcony or a door to an Alzheimer patient. Second, those related to the position of the person: if she is bended, squatted or even has fallen down and cannot stand up. Thanks to the 3D information with the location of the person the system can automatically detect any of these situations, which have been previously defined as risky.

Figure 3: The system has automatically detected the person faint.

When the system realizes that the person remains in a dangerous area or, for instance, is not able to stand up, then it automatically triggers the alarm and launches the corrective actions like sending a message or calling the human supervisor.

Using a PC for information processing makes it simple to send any alarm through Internet, or even to call to any Emergency Center through the regular mobile phone network. Such automatic alarm message may include an image or a video sequence, and so, a human operator may evaluate the situation and choose the right response to perform.

3. Visual 3D tracking technology

Several cameras are used simultaneously, pointed at the human user, in order to extract truly 3D information. Using more than one camera allows softer tracking and to extract depth information, reducing the computing power required to do so. The right number of cameras depends on the size and shape of the working area. All the relevant zones, at least the most important ones, must be covered by the visual scope of two or more cameras. Without it, the system will not be able to extract precise information about the user. Other data like when the person was seen for the last time is also available.

The image processing carried out by the system is based on the idea that making 3D hypothesis and checking them against the visual images is faster than building such 3D description from the images themselves. In our case, we use a particle filter to create and search for the best 3D hypothesis about the location of the person [2, 4, 5]. Once they are generated the system collects all the camera data together and checks whether the current hypotheses match the real images or not.

Figure 4: 3D monitoring with 4 cameras.

Figure 4 shows the images from the four cameras watching our lab area. A cloud of small red points is also displayed in all the images. Each point represents a 3D hypothesis about the person position inside the room. Actually the points are placed in the real 3D scenario, the images show their projections into each camera. If a given point is properly located in the real space, that is, is placed at the person location, then it will project into the parts of the images where the person appears (occlusions excluded).

In order to test if the projection of a 3D particle lies inside the pixels of the person in the image we use a simple color filter and a motion filter around the 2D projection of the 3D particle. The system counts the number of the filtered pixels around the 2D projection, which basically estimates the likelihood of the person being at the position of the particle.

Figure 5 shows an example of several hypotheses about the 3D position of a pink ball, instead of a person. Although the filters are quite simple, their combination and their persistence in time provides with a very robust and discriminant compound filter.

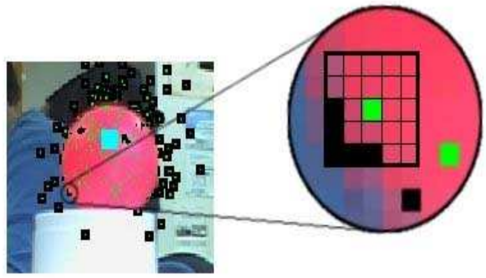

Figure 5: Color based observation model.

When the relevant color is not known a priori or not suitable, then the system could learn it automatically getting it from the color of the moving objects in the scene. As soon as an object is in motion in the working area, the system learns its colors and tunes the color filter to obtain it.

The combination of color and motion allows us to take advantage of both features. Motion let us to easily follow the person as the rest of the images in the scene remains static. In general, static objects won't be a target for our system. A motion detector will select most of the relevant data taken by our cameras but color data is also required since a person is not always moving and would otherwise be ignored by the system. Color filters help to easily discriminate relevant objects. The advantage of combining both features, motion and color, is that we can learn the color when an object is in motion and track such object when it doesn't move.

The ease of implementation of these filters allows checking tens of thousands of hypotheses per second in a conventional desktop PC. The algorithms used provide very high performance to the system.

4. Conclusions and future work

We have presented a system that automatically detects high-risk situations for elder people care and triggers the alarm when those situations arose. Specifically, it can detect a faint or fall. The alarm has an associated set of actions such as sending an SMS or a multimedia message with the video images to a mobile phone, calling a human supervisor, etc.

Our system has several video cameras, a computer and the required software for image processing. We use conventional hardware so that it can easily be replaced if needed and makes it a cheap system. The software uses advanced algorithms for three-dimensional images processing: particle filters.

We have built a prototype and tested it for fall detections in a room, triggering an acoustic alarm when people fall. The 3D algorithm used is very dynamic allowing an immediate detection of risky situations and a quick response to them.

Our system works without human intervention even if the person faints or loses consciousness. The user doesn't need to take any manual action or carry any device at all.

Compared to video care systems, where the images are sent to a remote location for an operator to keep the control, our system automatically detects the emergency. Privacy for the user is guaranteed and continuous monitoring is possible as it doesn't depend on anyone else.

At present we are working on several enhancements to our system increasing its functionality by monitoring several objects at the same time and detecting new risky situations.

References

[1] Vicon Peak http://www.vicon.com

[2] Barrera P., Cañas J. and Matellán V., "Visual object tracking in 3D with color based particle filter" , Int. Journal of Information Technology, Vol 2, Num 1, pp 61—65, 2005.

[3] Fritsch J., Kleinehagenbrock M., Lang S., Fink G. and Sagerer G. "Audiovisual person tracking with a mobile robot", Proceedings of Int. Conf. on Intelligent Autonomous Systems, pp 898—906, 2004.

[4] Pérez P., Vermaak J. and Blake A. "Data fusion for visual tracking with particles", Proceedings of IEEE, Vol 92, Num 3, pp 495—513, 2004.

[5] Pupilli M. and Calway A. "Real-Time Camera tracking using a particle filter", Proceedings of British Machine Vision Conference, pp 519—528, 2005.

[6] Zotkin D., Duraiswami R. and Davis L. "Multimodal 3D tracking and event detection via the particle filter", IEEE Workshop on detection and recognition of events in video, pp 20-27, 2001.