INTRODUCTION

Over the past years, the main ambition in telephony was to combine data, audio and video in a service called “triple play.” The idea is that the home will have a single point of entry for data, audio and video (or Internet, telephony and cable). This makes information a public utility in the same way as water, gas and electricity. From a historical review on the impact of such utilities on society, it can be easily concluded that the low-threshold access to information will shape the coming society (Carr, 2008).

The audio/video center has become a focal part of the modern household. In such a typical consumer system, the ingredients are the command console, the information & computation servers and the reproduction devices such as speaker and display. Meanwhile the wireless telephone has become the subject of convergence of camera,
music player and telephone into a single device, becoming an amusement center on its own (Pant, 2005). By virtue of its attractive support of mobility, this cordless has gradually replaced the wired telephone in the home, but the limited bandwidth of the radio-link has precluded a further merge with the audio/video center, which is therefore still handled by a separate remote control.

The early remote control devices simply send keyboard pressings over an infra-red channel to the TV or home director. This technique can be used for invoking a reaction from any object with an infra-red receiver on a distance limited to a couple of meters. A typical example is where museum visitors can hear an explanation on their phone when pointing at an art object (Pai et al., 2007).

Recently we see gaming computers added to the home amusement center. Games have been developed first as a PC application, gained momentum by dedicated consoles and gradually moved to use specialized peripherals. The wiring between computer peripherals and the processor box has always been a point of irritation. Cables got mixed up and the user was limited in its freedom. This certainly got more repressive with the coming of multi-user games. Therefore the need for wireless communication between peripheral and processor box came in demand.

So far only an information link is made by pointing and a command is sent. Then a reaction is given by the server on a display or a speaker. Meanwhile the motion sensor was developed in the automotive industry (Knivett, 2009). In a typical streak of disruption (Christensen, 1997), this device was then accepted in the digital camera to provide image stabilization, and introduced in 2007 by Nokia to the camera phone for similar purposes.

The real breakthrough of the motion sensor has come in the gaming industry. The movement of the console can be copied into the computer over a radio-link and interpreted to influence the game. Best known as Wii technology, it allows eliminating dedicated infra-red sensors on the objects and dedicated buttons on the pointing device by remotely moving a cursor on the display to the desired virtual button in a graphical user interface.

All such pointing and interaction functions can also be accommodated on a camera phone (Ludlow, 2008). The high computational requirements for processing increasingly large images seem to imply the use of mechanical motion sensors. But such high requirements are not applicable in case of interactive graphical user interfaces (GUI), for example in gaming. Where the GUI is local, the motion sensor has still clear advantages, as in the case of handling the screen on the Apple iPhone. But for cloud computing where the screen maybe remote, this sensor is not in the man-machine interaction loop. This leaves the question whether sufficient precision to capture just user directives (Cravotta, 2007) through image processing can be reached with limited local CPU load.

The paper is composed as follows. First we discuss different ways to capture movement, defined as motion plus direction. Then the basic block-matching technique is discussed and subsequently the extraction of movement by a limited search on images taken by a camera phone is treated. Then we provide some experimentation details and show room for further optimization. Finally we give a system’s perspective and draw some conclusions.

**CATCHING MOVEMENT**

The principle arrangement of remotely catching user directives is by watching him move. This usually requires a number of fixed cameras that each sent images to a central server who in turn creates and interprets a 3-dimensional picture. Calibration demands a well-prepared environment, but then a precise copying of the human moves is possible. A typical application is remote surgery (de la Barre, 2006) and community signage (Hildreth & MacDougall, 2009).