# THE HUMAN SENSING REVOLUTION

R. Perlovitch

Pebbles Interfaces, Israel

## ABSTRACT

Human sensing is the science of detecting human presence, count, location, posture, movement, identity, and even behaviour from sensory data. The need to sense people becomes ever more pressing, as the information is increasingly used to make decisions and provide services in diverse areas. However, due to technological constraints, human sensing have not been applied yet in the field of television, but there is no doubt that once implemented – it would mean a real revolution, no less.

Pebbles Interfaces managed to overcome these traditional technological challenges, and developed a new technology that finally allows a new era of human machine interaction for mass market distribution.

This paper presents an overview of some of the potential new innovative capabilities which televisions would have, using Pebbles' unique human sensing platform, and how it's expected to transform the industry.

## INTRODUCTION

In the last decade we witness a proliferation of devices with advanced sensing capabilities that measure motion, orientation, various environmental conditions, and most recently – human sensing. Human sensing is the extraction of information about the people present in an environment. The data coming from the sensor provide indication about people presence, count, location, tracking, identity, and behaviour. Using this data questions such as 'How many people are in the room?', 'What is their position?', 'What is each person doing?', 'Are they comfortable?' and so on can be answered.

Such human sensory information is used in many types of big data applications in various industries for better decisions making and to provide advanced services. The applications range from simple implementation like open a door as people pass or lock a computer when the user goes away, to a more sophisticated usage such as biometrics-based user's authentication.

Innovative human sensing has been adopted already in areas like security, medical, robotics and artificial intelligence, to name a few. However, at this point of time, due to technological constraints, the usage is limited mainly to research and enterprises, rather than home or personal usage. Therefore, human sensing have not been applied yet in the field of media (television, digital signage, etc.) despite the tremendous value it is likely to bring to this market.

Once human sensing capabilities could be introduced into the media market it would revolutionize this market and transform it from HW-based arena, to internet-dominate industry. Targeted advertising, personal content delivbery systems and interactive engagement will be possible, offering an entirely new viewing experience.

In this paper we review the traditional approaches for human sensing and explain their limitations. Then, we introduce Pebbles Interfaces unique technology, which enables for the first time, to break through existing barriers and allow a new era of human-machine interaction. Afterwards, we provide a glimpse to the new world of opportunities that will be possible once Pebbles' technology is adopted, and how it will transform the media market, and TV in particular. A discussion of possible directions the market could evolve into is given subsequently

## TRADITIONAL HUMAN SENSING

When considering human sensing for the media market, there is a number of key parameters that should be taken into account. These parameters include robust real-time high-resolution sensing capabilities within a certain range and field of view, which allow intuitive user interaction. Low cost and small dimensions are also important. Therefore, only optical technologies are relevant to this discussion. Other technologies, which address just part of the human sensing requirements are not discussed in this paper.

Traditional optical human sensing optical based solutions include 2D RGB images / video and general depth mapping.

### 2D RGB Images / video

Compared to other sensors, cameras are affordable, and a great number of devices have high-quality camera by default. Thus, camera-based human sensing usually requires a software-only solution and this field of computer-vision is highly developed. However, person-detection or identification of person-related activities is a great challenge using 2D images.

There are several reasons for this challenge. The first is quite obvious – The world is three-dimensional and not two-dimensional, which means that understanding of 3D structures and z-axis movements from regular 2D image is not a straightforward task. It may be doable but will require massive computing resources. In addition, there are many other fundamental challenges of 2D RGB images: it is difficult to distinguish between objects with a similar tone, movement may lead to smearing which impede the processing, and finally, lack of lighting causes data loss.

It is often advantageous to use depth information from stereo cameras as an additional cue to differentiate people from the background scenery. But of course this method will have same issues as described above.

### General 3D projection

Depth data provides essential information which cannot be extracted from 2D data. 3D projection is any method of mapping three-dimensional geometric model to a two-dimensional plane, usually in terms of point cloud (a set of data points defined by X, Y, and Z coordinate) that is later post-processed.

Common methods project an electromagnetic wave, typically in visible or infrared spectra, onto the scene, and measure the changes in specific properties of the reflected waves, which are then mapped to geometric quantities.

Time of Flight (ToF) and Structured Light are both well-known 3D depth sensing methods. ToF sensors operate by emitting modulated infrared light and measuring the phase shift of the reflected signal. Structured light is using a projected pattern that is observed by a camera with a known baseline distance.

In principle, ToF and structured light technologies have great potential for human sensing, but the inherent tradeoff they have between resolution and cost limit their use. They either used for low resolution full body applications, such as entertainment or for applications in which price is not a limiting factor, like military applications. For mass market usage, which require a combination of high resolution at low price (derived from low computation) and small size, general 3D mapping technologies are not viable, as they are incapable of detecting small objects at far distances at a reasonable system size and cost.

In ToF technology, each pixel requires its own processing unit, which means the pixel size is large, and to achieve high resolution a lot of pixels are required, resulting in a large and expensive system. In structured light technology, to achieve high resolution, dense light pattern is a must, which makes the optical system complex and difficult to design and produce in mass manufacturing. In addition, dense pattern at a large distance results in overlap of the pattern at short distance, meaning the system cannot provide high quality data in a large range. The heavy processing also dictates the use of a dedicated processing chip, resulting in high-cost solution. Finally, systems which rely on high power processing unit cannot meet low power criteria, which is essential in devices for mass market.

## BREAKING THE CEILING GLASS OF HUMAN SENSING FOR MASS MARKETS

Pebbles Interfaces introduces a new paradigm for human sensing, based on its unique optical segmentation capabilities. Figure 1 demonstrates the different human sensing methodologies, starting from the basic method that relies on 2D RGB camera, through general 3D mapping, up to a tailored approach of optical segmentation which offers the most advanced capabilities.

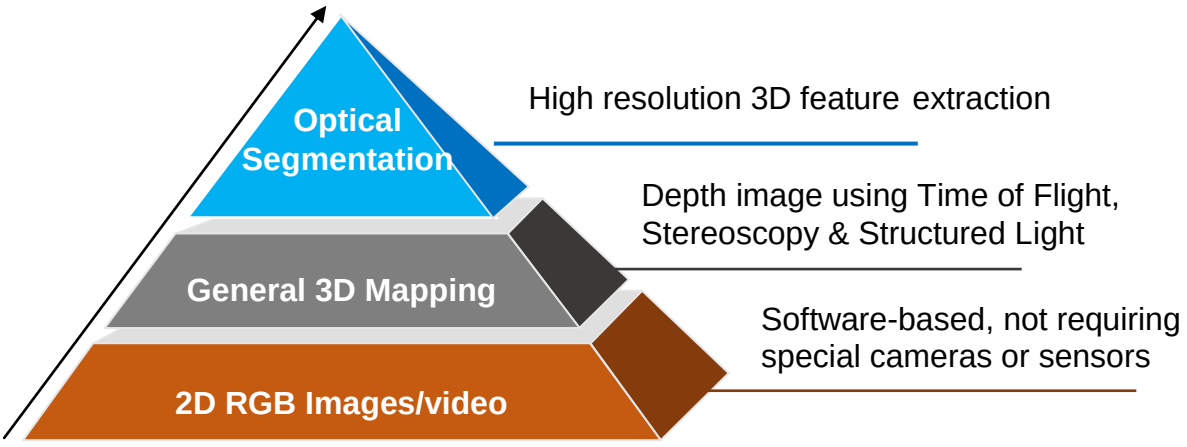
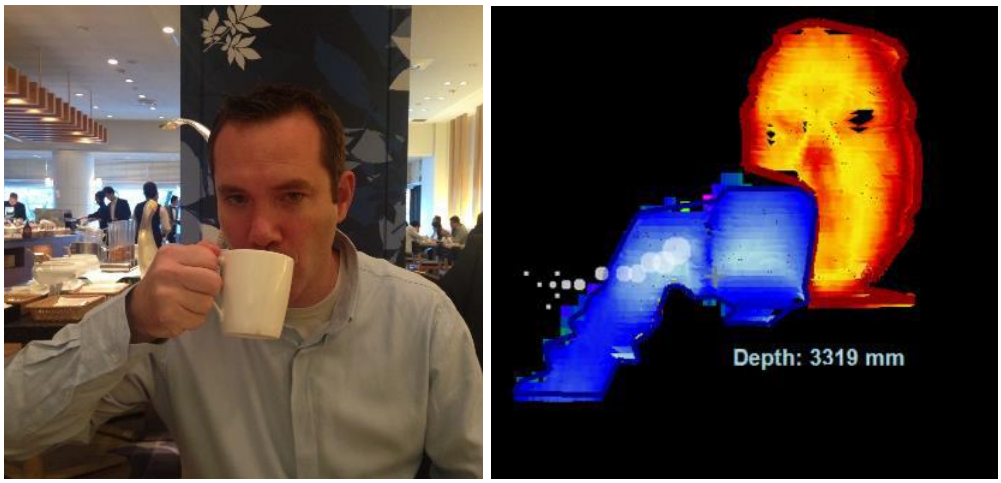



Figure 1 – Optical human sensing pyramid of methods

Pebbles Interfaces' optical segmentation allows direct segmentation of human body parts and absolute background removal even before generating a depth map (see Figure 2). Subsequently, a selective high resolution 3D depth map for objects of interest is created. Meaning, the analysis is at a local level, allowing to reduce the computational efforts to a minimum.

The system is based on proprietary projected Near-IR light pattern and off-the-shelf IR camera. The projected light pattern is encoded in a sophisticated way which ultimately provides highly sensitive human sensing, including head and body parts down to finger tips, nose, eye sockets and body curves.




(b) Real-life RGB image

(a) Pebbles Interface 3D sensor raw data

Figure 2 – Depth based accurate segmentation of human body parts

As stated, with general 3D mapping, a 3D depth map of the entire scenery is generated as a preliminary stage, which is a very heavy computational procedure, and only then, based on the 3D image, features will be extracted and background noise will be eliminated through image processing.

Pebbles Interfaces' radically different approach allows a solution which is both very sensitive as well as lean in computation, thus, cost effective, making it an ideal platform for human sensing applications for the mass market.

## TECHNOLOGICAL NEW APPROACH

The enabling factor is the usage of optical segmentation which as aforementioned, allows direct segmentation of human body parts and absolute background removal even before generating a depth map. This is achieved by encoding IR pattern in pseudo continues light features which provides local understanding of the examined object. Once an object is found to be relevant, depth map is extracted with local depth analysis in HD.

The pattern enables smart sampling in order to minimize the scanned pixels of each image which is significantly lower than the actual image size.

As the object is already segmented at the end of the process (finger tips, palm, etc.), the next layer is the application layer directly without additional computation of the depth map.

The system can be embedded in a device or operate as a standalone product connected to a device by USB interface. Since the computation is relatively low, the computation engine can be executed in most cases on existing application processor, with various operating systems (Android, Windows, etc.), and the need for a separate dedicated processing is eliminated.

Lastly, the fact Pebbles Interfaces can determine the depth of an object locally without extracting a larger area around it enables to use the resolution of the camera sensor to its full extent. Therefore, a very high resolution can be achieved while maintaining low computational requirements. Moreover, the higher the resolution of the IR camera, the more sensitive the detection would be, without a need to replace the transmitter itself.

## A NEW WORLD OF OPPORTUNITIES

When using Pebbles Interfaces' core technology of diffractive optics and detection algorithm, high quality visual information is available and enables advanced processing and insights about the people within the detection range.

The outcome, is a new engagement and interaction platform, which offers different levels of information with various possible capabilities:

| Detection Layer | Details |
|---|---|
| Human Presence | Human presence identification, count, age/sex classification |
| Body parts | Segmentation and tracking of body parts (head, torso, hands including finger tips, down to fine body elements like nose, eye sockets and body curves) |
| User ID | Distinguish between anonymized people, identify registered users |
| Posture | Analysis of body parts' position, correlation between different body parts, and facial expressions |
| Behaviour Detection | User engagement analysis based on integration of all related visual properties (head position, head nodding, leaning posture, hand position, etc.) according to a defined involvement scale<br>Social connections classification based on a database |
| Interaction with objects | Identification of objects (based on predefined list) and type of interaction with the objects |

Table 1 – Detection layers and related information

The new capabilities practically give system the ability to see and understand their environment, extending the capabilities of artificial intelligence in way that could be relevant for mass market.

Below are examples of the detection layers listed above in the “eyes” of Pebbles Interfaces system:

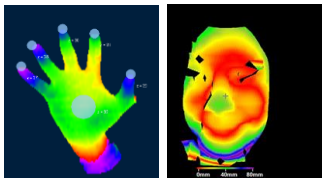
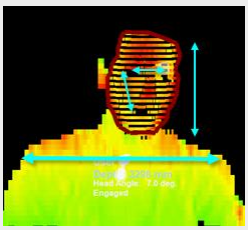
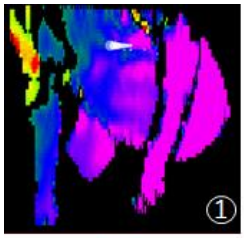
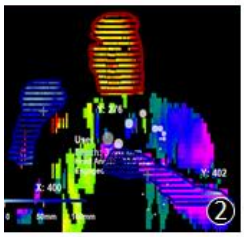
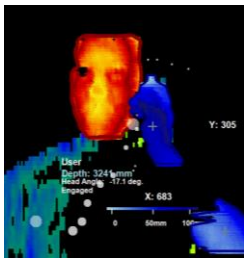
| Detection Layer | Visual Example | Example for Related Output |
|---|---|---|
| Body parts |  | • Hand and Face segmentation & tracking<br>• Finger tips & center of mass location<br>• Tracing body curves and orientation |
| User ID |  | • Shoulders width<br>• Head’s dimensions<br>• Distance between the eyes<br>• Angular distance between the edge of the eye and the tip of the mouth |
| Posture |  <br>  | 1. Chest forward, neutral arms posture (may suggest user is sitting down)<br>2. Fisted hands, moving back & forth (implies on punches) |
| Interaction with objects |  | • User is holding an object<br>• The object is classified as a phone<br>• User is talking over the phone |

Table 2 – Examples of detection layers and related information

When focusing on the television market, for instance, we can demonstrate a few completely new capabilities, powered by Pebbles Interfaces’ human sensing. Later, we will discuss the great value it can bring to the industry, and transform it from HW-based into web-based.

**New natural and intuitive control system**

Thanks to the high sensitivity of the system, even at far range (up to ~5 meters), users can use subtle hand and finger gestures, including multi-fingers gesture, while at ease, to control their TV sets remotely.

This lean-back experience, supported by a huge variety of possible gestures, includes the following representative capabilities (Figure 3):

- Point, Click, Select – point at a spots/area of interest,
- Rotate – realistic twisting, feature extraction
- Hold, Move – bring closer/further, move around, zoom, drag, pinch, swipe
- Social actions – thumbs up/down, sign language, play

Figure 3 – Examples of natural and intuitive hand and finger gestures

**A unique system for identifying the level of engagement**

Pebbles Interfaces technology enables real-time continues tracking of the vertical and horizontal angles of the user's head and upper body. Therefore, user engagement can be identified instantly by detecting whether its position is directed towards the TV (engagement can be defined by a certain range of angles).

Additional information can be extracted, such as the total engagement duration, or synchronized the engagement level with time / content delivered at that time. When the user's head fall down or sideways it may suggest the user is asleep.

Figure 4 outlines several examples of engagement modes which can be detected by Pebbles interfaces technology (direct sensor output, not post processing).

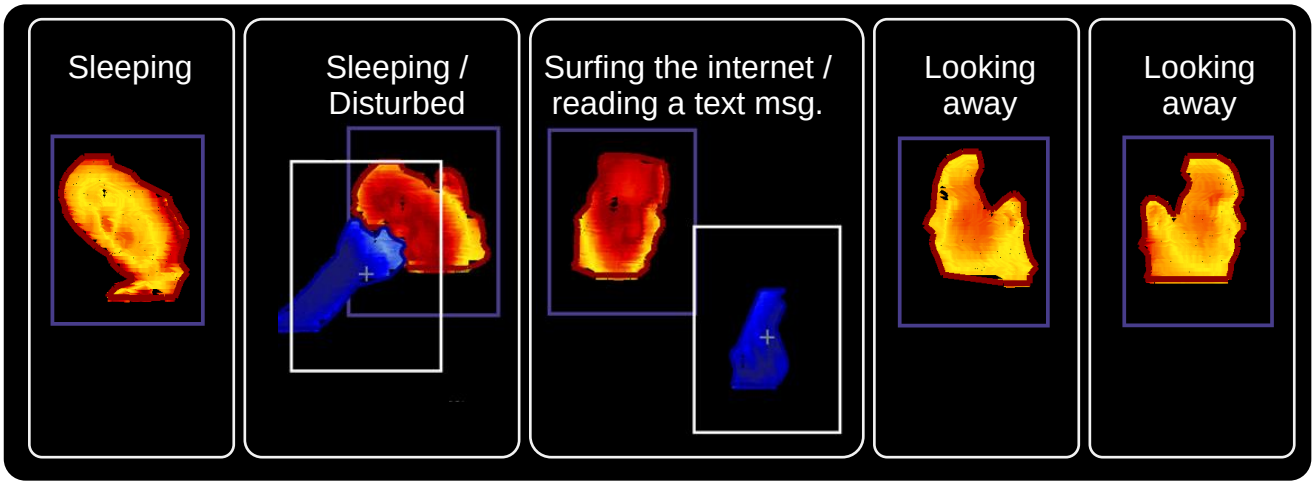



Figure 4 – Examples of engagement modes detected by Pebbles Interfaces optical segmentation

## TRANSFORMING THE TV MARKET

The TV hardware business has cooled off since the heyday of big flat-screen. The content business is hotter than ever, with more providers offering more choices of programming to more types of devices so the TV business is no longer just for televisions.

Nevertheless, exciting business opportunities such as targeted advertising, personal content (that is not based only on the identity of the viewer, but also on his/her level of concentration and body language), and accurate real-time rating measurements are not exploited at all.

The TV is still considered a “dumm” device which is insensitive to the people who are using it. It does not take active actions, does not recognize its users, even if they use it regularly. It doesn’t know their preferences, unless they proactively show it, and it does not respond to what is happening while it’s on.

Pebbles interfaces’ disruptive human sensing technology can change that.

Pebbles Interfaces would provide televisions, for the first time, the ability to understand situations and customize its “behaviour” accordingly.

For example, if the person who is watching the TV gets a phone call and looks away from the TV screen for some time – the TV will mute itself, or automatically pause the program. When that person finishes the call, the TV could offer him/her to continue watching from where they stopped.

Content could be adjusted to the user who is watching. For instance, it would be possible to limit the content kids view. If an adult is watching violent content and a child steps into the room, the screen can be darken, or the TV can switch to a different channel. Parents could configure what channels their children are allowed to watch, and of course, if there are children of different ages, for each one a complete different definitions can be set.

Instead of commercials that are broadcasted according to accumulated statistical information, advertisements could be tailored to the individual user in real time, maximizing the commercials’ impact. No more broadcasting stereotypes ads such as feminine hygiene products during morning shows, and beer or cars during a soccer game, without matching it to the actual viewer. Ads could be much more relevant and impactful when tailored for the person who is watching them. For instance, if the viewer is a man who is drinking from a cup while watching TV, it may be a good time to broadcast a commercial of new cookies which are a great addition next to coffee or tea. That person may be more easily influenced by this commercial than at any other time, because he is now in the right state of mind. Without human sensing there is no way to have this type of information.

The aforementioned examples provide only a partial glimpse of the many more possibilities that could be available once powerful human sensing technology, such as Pebbles Interfaces’ is adopted into the media market. A new world of opportunities will open up, the sky is only the limit.