Portable System for Reading Display Panels

F.J. González, J.A. Gutiérrez, A. Picón, A. Isasi, A. Domínguez, I. Idigoras
ROBOTIKER-TECNALIA
jgonzalez@robotiker.es

Summary

This article outlines the DISPLAYER project, the aim of which is to develop a system enabling the blind and partially sighted to read, using a mobile device, the different information screens (display panels) which are increasingly common on domestic appliances and everyday household equipment. Based on artificial vision technology, a portable system has been proposed which is designed to detect and interpret the visual information shown on the different types of display panels (numeric and alphanumeric characters and iconography) and transmit such information to the user via voice messages. The article describes the approach adopted to develop the system as well as the progress and results obtained so far.

1. Introduction

The majority of the population views intelligent homes as sophisticated devices which provide people with comfort, however, for disabled people this technology can mean a significant increase in their quality of life. Nevertheless, these devices require interfaces which can be used by people with impaired vision [14].

Currently, the amount of electronic equipment, electrical appliances, etc in the home which have display panels to provide the user with information is increasing. We come across these systems during our everyday activities on a more and more frequent basis: microwaves, audiovisual equipment, thermostats, boilers, ovens, hobs, digital clocks, fridges, supermarket checkouts, and many other devices around us have display panels. Making this information accessible to people with visual disabilities represents a significant benefit when it comes to the independence of this group of people. Being able to access the information provided through visual media (display panels) is of great importance to this group. Two routes for making this information accessible are being considered:

1. That the manufacturers of these devices themselves incorporate technology such as voice synthesis technology in their products which transmits the information presented visually on display panels by sound. Some examples of this approach are known to exist; however, they have resulted in expensive, top of the range products.
2. The second route is to provide technical aids which can read the aforementioned display panels and translate the visual image into sound.

It is this second route that we wish to explore in this project in which the development of a **portable device designed to detect and interpret the visual information shown on the different types of display panels (numeric and alphanumeric characters and iconography) is being considered. DISPLAYER is seen as a useful and affordable tool for the blind and people with visual disabilities generally, which enables them to independently access the dynamic information shown on display panels, allowing them to improve their quality of life by integrating in a more seamless fashion with the environment surrounding them**.

2. Approach and adopted methodology

2.1 Display panel types

The various everyday electronic appliances include a wide variety of models and different types of displays which may be classified as follows:

According to the technology:

1. LCD (*Liquid Crystal Display)*, made up of electronic elements which are transparent or opaque depending on the charge applied. This type distinguishes devices which have a: passive background (which reflect the light affecting the device using a reflective background surface and the characters appear dark) and an active background (illuminated). Figure 1 shows two examples of numeric display, one passive and the other active.

- LED (*Light Emitting Diode)*, made up of electronic elements which emit light when a charge is applied to them. At the bottom of Figure 1 there is an example of a numeric LED display which shows the time.
- TFT. *“Thin film transistor”* A colour graphics screen, the active elements (pixels) of which permit the subsequent modification of the colour of active light. They display high resolution images.

Figure 1: left to right, top to bottom, passive background LCD (dark elements on a reflective background), active background (illuminated) and LED (light emitting elements).

According to how the information is displayed:

- 7 segments. The display elements are 7 bars. Numerical and some characters.
- 16 segments. Permits alphanumerical display.
- Dot matrix. Rows and columns of characters created by dots (usually 8x5).
- Graphics. Dot matrix (pixel). Characters and configurable graphics. (320x200,…etc)

Figure 2 shows a few examples of the different types of display.

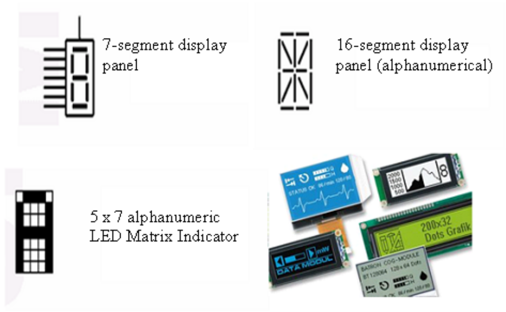

Figure 2: left to right, top to bottom, 7 segment, 16 segment, dot matrix and graphics display panels.

2.2 Technology status

In the university area, mobile character recognition systems (OCRs) are being developed, specifically aimed at people with visual disabilities, as well as display panel readers like that shown in Figure 3 which contains a screenshot displaying the image captured from a video which shows the time and possibly the selected channel number [4, 5, 6, 8, 9].

Figure 3: Example of display reader development .

It is a 7-segment display reader computer application [5] with its own lighting. This prototype is based on a simple light-dark contrast taking into account the intensity of the level of grey. However, this technique does not allow many current display types to be adapted, such as colour LCD screens or display panels with little contrast. This solution focuses on analysing numbers and letters not being able to interpret other more complex types of information, such as icons or indicator bars.

Another research prototype bases the system on a special PDA [4], which has a camera and an embossed keypad to make it easier for the user to operate. The research group has observed the issues faced by people with visual disabilities when it comes to taking photographs, for which they have designed a standard card using which they can embark on photography training supervised by the system itself. In 2004, the university reported a 33% success rate in reading characters using commercial OCRs, which rises to

69% using their methodology. In 2005, they launched a system based on neuronal networks, which aims to correct the identification of text taken in bad conditions increasing the success rate to 88%.

In the meantime, applications using the continually improving image capture and processing capabilities of mobile telephones to interpret codes, symbols and other visual images are starting to appear.

2.3 Adopted approach

In this context and given the large variety of display panels in existence, as well as the types of information they display, the proposed methodology for extracting the information from images is as follows:

1. Extraction of image features and identifying elements of interest.
2. Identification of symbolic information in that image.
3. Intelligent processing of the image based on the semantic knowledge of what can be seen.
4. Coherent communication of the information compiled.

An important aspect which has been considered from the start of the project is that of images which are captured in conditions which are less than ideal, as this action will be performed by people with visual disabilities.

The development requires the captured image to be adapted and interpreted by the system, in a similar fashion to that performed by human beings. The system must be capable of extracting not only the identity of the various parts making up the image, from a set of numerical data, but also understanding its content, so as to be able to give the user the requested information.

3. Development status

Currently points 1 and 2 of the aforementioned methodology are being tackled which include image capture, the development of algorithms to identify and locate the elements of interest, and their interpretation. The results obtained so far are presented below.

3.1 Image capture

It is anticipated that the system will be based on consumer electronics. In this sense, hardware development has not been considered, however an image capture methodology using various sources has been carried out, enabling an initial comparison of the differences in image capture between different media and users during this first stage.

In order to achieve this, image capture tests have been carried out using at least four different devices (a webcam, a mobile telephone and two low-resolution, low-cost digital cameras). Although in the final project, the processing system will be based on a mobile device, for the development phase an architecture based on a laptop has been relied on, which enables a more straightforward software integration. In the next phase of the project the algorithms will be migrated to compact systems. Figure 4 below shows the platform composed of the laptop and, in this instance, a webcam.

Figure 4. Image capture and processing devices used (laptop and camera).

Likewise an image capture methodology has been established using which, by issuing a few basic image capture guidelines, an extensive series of captured images has been collected by over 80 people using 500 appliances in their homes in conditions close to visual disability (image capture by feeling the device to ascertain the position of the camera and ambient light). This enables the control of variables such as:

- Display differences
- Lighting differences
- Orientation differences
- Capture differences

Figure 5 shows some examples of the captured images which come from two microwave models. In the images the display panels show icons, numbers (indication of the time and the programme) and text (“end”).

Figure 5. Examples of captured display panel images.

3.2 Extraction of image features and identification of elements of interest.

Firstly, it is necessary to extract image features. It is necessary to segment the image into different areas, in other words, to separate the relevant information from that which is not, contained in the element being analysed, in this case, the display panel. In order to achieve this, segmentation algorithms based on texture (invariant Gabor filters) [11, 12] are used, allowing us to divide the image into various sub-images, according to their texture. Only the areas whose texture responds best to certain frequencies (display panels) will remain.

Figure 6 shows an example of the application of segmentation algorithms based on textures. It includes the starting image of an alarm clock and the image following the application of these algorithms in which all the digits and areas of interest on the display panel can be clearly distinguished.

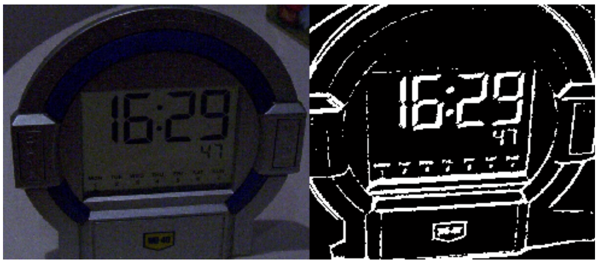

Figure 6. Example of extraction of elements of interest based on texture characteristics.

Figure 7 shows a three-dimensional diagram of one of the filters used in this segmentation.

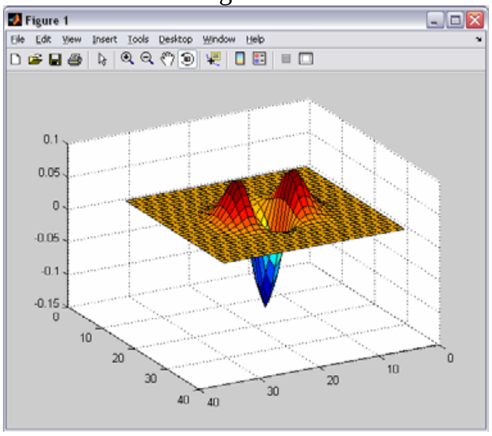

Figure 7. Three-dimensional diagram of one of the filters used in segmentation.

As Figure 8 shows, by applying classic binary techniques to the same image the quality of the results of the treatment are appreciably inferior, causing some areas of interest to be lost and a less defined extraction of the actual display panel.

Figure 8. Extraction of elements of interest, using classic binary techniques.

3.3 Identification of symbolic information in this image.

The elements of interest are analysed and grouped according to common characteristics of position, size, orientation and texture in various blocks making up “similar elements”. These are classified as textual or non-textual elements, and the interpretation and reading of the content is performed based on the result.

Once the image has been separated into several sub-images, it is necessary to interpret each one of them [10] (what does the text say, what is the number worth, what does the icon mean...).

The classification of the sub-elements of the display panel, based on the features extracted from them, is tackled using pattern recognition algorithms [13]. These algorithms allow us to establish the relationship between the set of pixels in an image (with no obvious relationship between them) and the interpretation that the human brain makes of the image.

There is a very specific set of pattern classification algorithms for identifying written text: OCRs [1, 2, 7] (Optical Character Recognition). These systems resolve the issue of one letter being represented in many different way (different font types) and, nevertheless, still refer to the same letter. These algorithms are considered to be the most suitable for correctly classifying and reading the parts of a display panel which are classified as text/digits. However, the recognition of graphic icons, volume and status indicators, units... is tackled in the next phase, from a standpoint based on the generic classifiers mentioned above.

Below, in Figures 9 to 11 there are several examples of display panel image processing, in this instance of a video recorder on which text (“play” and “stop”) and digits with symbols (“—12”) appear. In the images the symbols recognised by the system are enclosed in yellow boxes.

Figure 9. DVD display image and its processing.

One of the limitations of the current OCR algorithms is that the text cannot be rotated, should not have variations in perspective or have excessive segmentation. As a result, it is necessary to reprocess textual elements to make them suitable for display panel recognition. This involves reorientating the identified element and correcting its perspective, rotation and scale, as well as appropriate segmentation and adapting the elements so that they can be interpreted by the OCR system. Figure 10 shows the results of processing a rotated image from a video recorder display panel containing the word “stop” through the system.

Figure 10. Display image and its processing. The image rotation is significant.

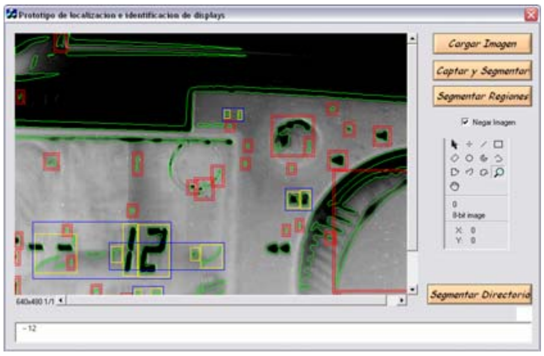

Figure 11. Video display image and its processing.

4. Tests and results

The prototype in the first phase of development includes the following features:

1. Automatic image capture.
2. Extraction of image characteristics and identification of elements of interest. Improved image quality due to preprocessing (segmentation,

rotation, perspective) and an advanced adaptive filter.
3. Identification of symbolic information in textual elements using OCR
4. Audio conversion and transmission to Bluetooth headset of system response.

The system has shown itself to be fully operational with 7 and 16-segment display panels for an initial range of selected devices focussing on entertainment systems (video, DVD, sound equipment). With close to 92% accuracy.

Furthermore, tests have been carried out on the simple images captured from a wide range of devices demonstrating extremely high levels of accuracy.

5. Conclusions and future work.

The results obtained so far allow us to conclude that:

- The system is capable of resolving the issue of reading 7 and 16-segment display panels and enables us to foresee further progress.
- In the medium term, this development can become a highly powerful too to increase the personal autonomy of those people with a visual disability.

The next steps to be tackled on the project include:

- Image capture by people with a visual disability using a wide range of devices in real conditions so as to verify the methodology and have another set of images.
- Improvements to image characteristic extraction, element of interest identification and symbolic information identification algorithms. Incorporation of icon recognition.
- Intelligent image interpretation based on semantic knowledge of all the elements detected and classified by the system.
- Migration of algorithms to a compact PDA-type device or smartphone .
- Communication of the interpreted information via voice synthesis technology (TTS).
- Compact device testing by end users.

6. References

[1] "Optical Character Recognition. Using Onmipage", http://help.phys.unsw.edu.au , June 2006.

[2] Instructional Technology Resource Center (ITRC), http://www.isu.edu/itrc, June 2006.

[3] Mediaseek Inc. "CamReader". *http://www.camreader.jp* , June 2006.

[4] Silvio Ferreira, *Faculty of Engineering*, Mons (Belgium), http://www.tcts.fpms.ac.be, June 2006.

[5] H. Shen and J. Coughlan, "Reading LCD/LED Displays with a Camera Cell Phone", *2nd IEEE Workshop on Embedded Computer Vision (ECVW '06)*, in conjunction with *CVPR 2006*, New York, June 2006.

[6] New Interfaces. The vOICe. http://www.seeingwithsound.com/ , June 2006.

[7] WATA Bulletin 2001, "Omnipage", http://wata.org/pubs/bulletins/fall2001.htm, June 2006.

[8] "Aids for visual impairment" http://www.asphi.it/english/Tecnologie/DisabVisive.htm , June 2006.

[9] "Mobile OCR, Face and Object Recognition for the Blind", *The vOICe*, http://www.seeingwithsound.com/ocr.htm

[10] C. Hudelot and M. Thonnat, "A cognitive vision platform for automatic recognition of natural complex objects", *ICTAI*, Sacramento, USA, November 2003.

[11] Ojala, T. and M Pietikäinen, "Texture Classification", *Machine Vision and Media Processing Unit*, University of Oulu, Finland, January 2004 http://homepages.inf.ed.ac.uk

[12] Picón A. Et Al. "State of the art: Texture-based image classification and segmentation methods, 2006".

[13] Duda et Al, "Pattern Classification", *Willey-InterScience* 2000

[14] Gill, J., "Access-Ability, Making technology more useable by people with disabilities", 2004