Bionic eyeglass: an audio guide for visually impaired

Kristóf Karacs, Anna Lazar, Róbert Wagner, Dávid Bálya, Tamas Roska

Faculty of Information Technology, Pázmány University and
Analogic and Neural Computing Laboratory, SZTAKI, Hungarian Academy of Sciences

Mihály Szuhaj
Hungarian National Association of Blind and Visually Impaired People,

Budapest, Hungary

***Abstract*—In spite of the impressive advances related to retinal prostheses, there is no imminent promise to make them soon available with a realistic performance to help navigating blind persons. In our new project, we are designing a Bionic Eyeglass that provides a wearable TeraOps visual computing power to advise visually impaired people in their daily life. In this paper the system aspects are explained. Basic tasks are indoor and outdoor events, defined by blind people, categorized into different situations. Two types of cellular wave computing algorithms are used: general purpose spatial-temporal event detection by analogic subroutines developed so far, and recently developed multi-channel mammalian retinal model followed by a classifier. The basic idea is to mimic the way the nervous system discriminates relevant information from the irrelevant mass – namely realize an attention model. Typical indoor and outdoor event detection processes are considered.**

I. Introduction

In spite of the impressive advances related to retinal prostheses, there is no imminent promise to make them soon available with a realistic performance to help navigating blind or visually impaired persons in everyday needs. In our new project, we are designing a Bionic Eyeglass to give them support in their daily life. The presented system differs from existing topographic classification techniques in the intensive multi-channel retina-like preprocessing of the input flow, as well as the specific semantic embedding technique. The system is designed and implemented using the Cellular Wave Computing principle and the adaptive Cellular Nonlinear Network (CNN) Universal Machine architecture [1, 2, 3].

ONR N00014-02-1-0884 and
RET Szentágothai Tudásközpont 2005

There is a strong biological motivation behind building a multi-channel adaptive algorithmic framework. The CNN-UM proved to be a suitable tool for modeling the multichannel processing of the retina in which each of the channels extract different spatio-temporal features of the input. [4]

Our Bionic Eyeglass makes a major difference compared to any other devices made for visually impaired people since it is based on

- a cellular visual microprocessor family developed via the CNN Universal Machine principle with unprecedented computing power on a ~1 cm2 silicon chip with ~1W dissipated power,
- a dual visual input architecture (called the Bi-i [5]), and its software technology [6] and system implementation based on the above type of microprocessors,
- a multi-channel mammalian retinal model [7] based on the recently discovered retinal operation and implemented real-time on the Bi-i., and
- cellular wave computing algorithms combining topographic and non-topographic multimodal sensory flows [6].

A specific objective is to communicate the recognized objects and /or situations to the impaired persons by sound (speech). The research, design, and experimental implementation of the hardware and software tasks will be followed by the practical clinically supervised tests. In this paper the system aspects of the Bionic Eyeglass are explained. The next section outlines the system requirements, design and architecture. The third and fourth sections show the color processing and some details of the partially neuromorphic saliency and event recognition system.

II. System Design and Architecture

The Bionic Eyeglass provides a wearable TeraOps visual computing power to advise visually impaired people in their daily life. There are three different types of common situations: home, work, and on the way between them. A few standard image flows with some auditory information is used as benchmark. The basic tasks are indoor and outdoor events, defined by blind people.

Though we tried to restrict the task by selecting some typical places (home, street and office), the proposition is still very complex. The algorithmic development starts from the former analogic CNN algorithms for recognition of door handle and door sign [8], as well as object avoidance mechanisms [9], integrates the cellular wave computing algorithms for typical situations and blooms to a neuromorphic system with attention-selection and semantic embedding. The hardware implementation platform evolves from the present Bi-i self contained unit and mobile phone platform to a single, integrated eyeglass-mount unit using a SoC.

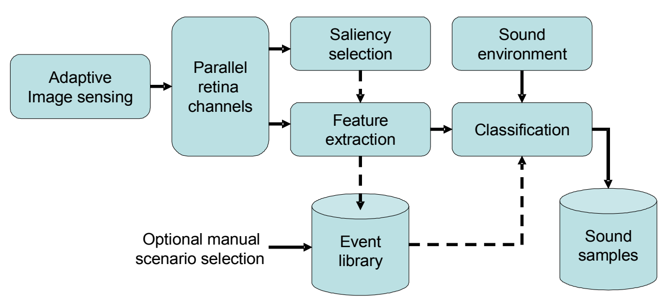


Figure 1. The partially neuromorphic system overview

Two types of cellular wave computing algorithms will be used: (i) stand-alone templates and subroutines and (ii) bio-inspired neuromorphic spatial-temporal event detection. Examples for the former one are the door handle detection, corridor sign extraction, bank-note and letter extraction [10]. The second type of algorithm is a neuromorphic saliency system [11] using the recently developed multi-channel mammalian retinal model [7] followed by a classifier using the semantic embedding principle (e.g. [12]). The system architecture is shown in Figure 1.

TABLE I TYPICAL TASKS CONSIDERED FOR THE BIONIC EYEGLASS

Place	Home	Street	Office
User-initiated functions	*Color and pattern recognition of clothes*	Recognition of marked and unmarked crosswalks	Recognition of control signs and displays in elevators
	Bank note recognition	*Escalator direction recognition*	Support in navigation in public offices and restrooms
		Public transport sign recognition	Identification of restroom signs
		Bus and tram stop identification	Recognition of signs on walkways
		Recognition of client displays (e.g: in banks)	
		Recognition of messages on ATMs	
Autonomous warnings	Light left switched on	Obstacles at head and chest level (branches, signs, devices attached to the wall, etc.)	
	Gas oven left turned on		

III. COLOR PROCESSING

Despite the fact that visually impaired people do not perceive color the information about the color of an object is important in cases like color of clothes, Figure 2. Thus we include a function that informs the user about the color texture of the objects seen.

In the planned system we will extract the colors seen on the scene and retrieve their location. The extraction of the colors can be understood as a color segmentation problem in which we segment the perceived image based on color information. Color segmentation algorithms have two main aspects, the first is the color space they use the second is the method they use to group the pixels in the chosen space [17].

The color spaces are linear or nonlinear transformations of the RGB color space. For the computation of the perceived color we use the nonlinear CIE Luv color space [18], which has the property that the distances of stimuli is similar to the human perceived chromatic distance. Hence the transformation of the RGB channels corresponds to the human perception.

The color specification algorithms found in the literature can be grouped as followings:

1) Pixel based classification. These algorithms classify the pixels based on their trichromatic values. They try to find groups of pixels in the 3D color histogram. The common techniques for this are thresholding or clustering.

2) Region based methods. They consist of growing regions from an initial seed until color boundaries are reached.

3) Edge based method. Similar to the region based methods. Here we specify the boundaries of the color objects, and then use region based method.

Steps of the algorithm are as follows (steps on CNN-UM are bold):

i. RGB → Luv
ii. Luminance adaptation
iii. Clustering
iv. Merging similar clusters
v. White correction
vi. Location color names

The method used for the color classification is the well known k-means clustering [19]. The color segmentation is only efficient, if we can get rid of the distorting effects of illumination. Hence we performed a local luminance adaptation method in which we eliminated the differences in the spatial low-pass component of the luminance of the perceived image similar to [13] (see Fig. 2 a. b.). Since the "L" channel of the Luv space represents the luminance we performed our local luminance adaptation on it. The great advantage of the CNN-UM architecture comes with local processing. This enables us to easily compute the local average for the luminance adaptation.

The clustering has the problem that it can produce an oversegmentation, because we do not know the proper number of initial cluster centers. We overcame this problem with a postprocessing step, in which we merged similar clusters, whose center lies close to each other. The result of clustering and merging can be seen on Fig. 2.

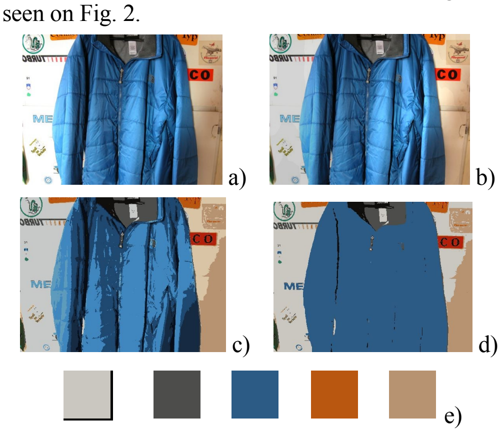
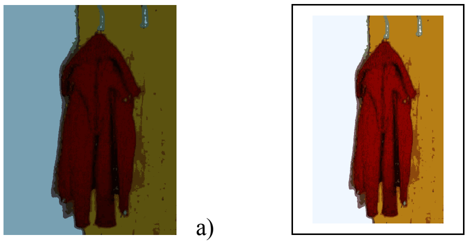


Figure 2. Clustering and merging of clusters. a.) shows the original picture taken by a standard digital camera. b.) shows the effect of the luminance adaptation. On c.) we can see the result of clustering. On this picture the pixels have the color of the cluster they assigned to. Regions of a given color represent a cluster. d.) shows the clusters after merging of similar clusters. The main cluster colors can be seen on e.)

The postprocessing step "White Correction" (see steps of the algorthim) exploits the fact that we know we have clustered the image. If there is a cluster that lies close to saturated white color we amplify the channels of this cluster to become white. The other clusters channels are also scaled with the same amplification factors. This correction can remove chromatic distortion of the images (see Figure. 3).

The last stage determines the location of the clusters and gives a verbal classification of their location the categories are: middle, left, right, sides (left and right), top, bottom, top-bottom.

The texture of clothes is also to be determined and classified in the following categories: uniform, striped, checkered or polka-ed.



a) b)

Figure 3. a) shows image taken by a mobile phone, which has extra sensitivity for blue colors. b) Shows the corrected clusters, we can see the white color of the wall.

IV. Some Details of the neuromorphic saliency and event recognition system

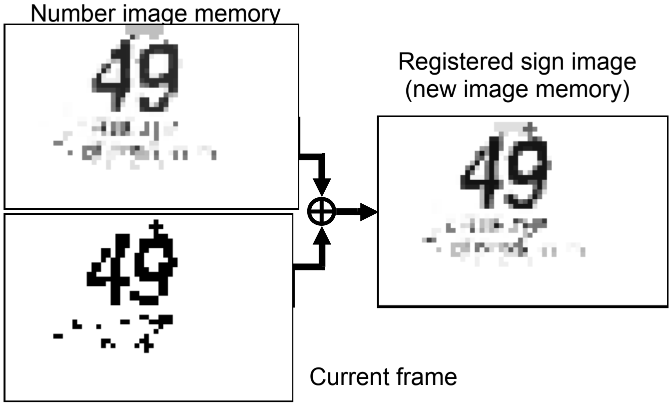
A. *Adaptive image sensing*

Adaptive image sensing is important if we deal with scenes that have large intra-scene dynamic range, like in real-world street image flows. Recent works [13] on adaptive image sensing using CNN-UM are developed using locally adaptable sensor array. A retina-like adaptation can be achieved by adjusting the integration time so, that the local average of an image region becomes the half of the maximum value. This eliminates the intra scene DC differences. In outdoor scenes where the variations of illumination might be large – both in time and in space – the adaptation is a useful property that enables the operation of the recognition steps.

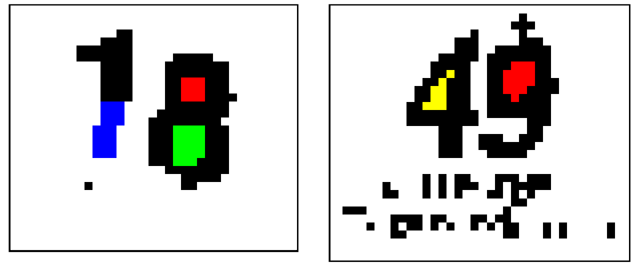
B. *Parallel image sensing- processing*

The first and best-known part of the visual system is the retina, which is a sophisticated feature preprocessor with a continuous input and several parallel output channels [14]. These interacting channels represent the visual scene by extracting several features. These features are filtered and considered as components of a vector that is classified.

Beyond reflecting the biological motivations, our main goal was to create an efficient algorithmic framework for real-life experiments, thus the enhanced image flow is analyzed via temporal, spatial and spatio-temporal processing channels. The outputs of these sub-channels are then combined in a programmable configuration to form new channel responses.



Enhancement of the number image by superposing the actual frame and the image memory.



Feature maps of route numbers: vertical line (blue), upper hole (red), lower hole (green), triangular hole (yellow)

An example for this is recognition of signs of public transport vehicles. This is a controlled situation thus the user can activate this function. The processing uses subsequent frames of the input video flow to recognize the sign. Algorithms modeling the channels first locate the sign on the scene, then extract features and classify the number (see [12] and [21]). The process is shown on Fig. 1 and 2.

Due to the low resolution of the images and the high level of noise present on them the binary number images become vague. To overcome this problem we make use of the a priori knowledge that the signs normally do not change, which means we can superpose subsequent sign images to achieve better image quality.

For number recognition we use topographic shape features that can be extracted by cellular wave algorithms. In the first step the number of figures in the number is determined by counting connected objects on the image that are bigger than a threshold. Features used include holes and lines. Holes are classified based on shape, size and position, whereas lines are classified according to orientation (horizontal or vertical) and position. Algorithmic details are described in [21].

C. *Saliency selection*

Visual attention is an ability to direct our gaze rapidly towards the objects of interest. This is a very complex mechanism, which includes two different, but tightly together, parallel working methods. These are the bottom-up (or image-based) and the top down (or task-driven) methods [11]. Bottom-up originates at the retina and goes towards higher brain areas (involuntary), while top down originates in the high brain areas and projects towards the muscles of the eyes (voluntary). We know much more about the bottom-up method, which basically filters out the salient, conspicuous, sudden and unexpected parts of the visual scene.

In nature, saliency is “calculated” with receptive fields (RF), where neurons are organized into concentric circles: a central- and a peripheral part that respond antagonistically. [22]

The flow diagram of the bottom-up process is depicted in Figure 4. In the first step the incoming vision is dissolved into several parallel retina channels, which are topographic maps of the visual scene. These channels code different low-level visual features, like motion, edges, colour antagonisms etc. In our model we use real retina channel emulations.

Once these channels are drawn up, each creates its own saliency map, which indicates that how salient, how 'loud' the different points are according to the appropriate low-level visual feature. These are also topographic maps, like the final (or master) saliency map, which is produced by aggregating the former ones and the most salient point wins the attention. The weights of the different retina channels are not the same, they change according to the actual tasks.

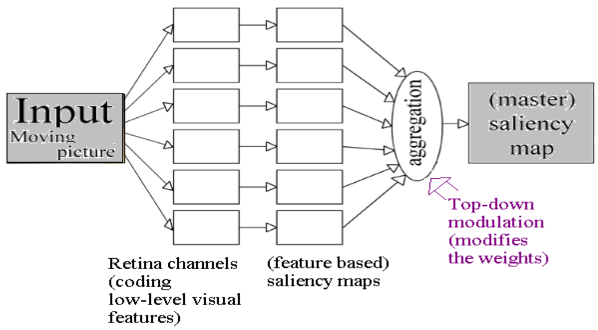


Figure 4. The flow diagram of the bottom-up attention mechanism.

One of the outdoor tasks that we would like to perform is to define the direction of the escalator. This is particularly important in those cases, when nobody or very few people is on the spot, so the blind person can not move with the crowd or can not ask. Figure 5 shows a potential solution for this task: looking for horizontal lines that can be filtered out from most of the retina channels. Even so we are using all of them –it would be unnecessary–: the transient (third picture in the first row), the local edge detector (beneath the transient) and the intensity channels are enough. If the detected bar moves upwards, i.e. its vertical co-ordinate lessens, then the escalator shoves out, otherwise it draws near. If the bars are steady, then the device is out of order.

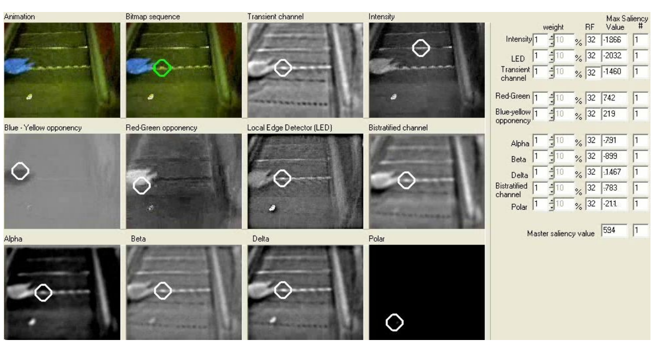


Figure 5. With horizontal line shaped receptive fields the edges of the staires will be filtered out. From here we only have to define weather the vertical cooridinates of these points lessen (the escalator draws near) or grow (shoves out).

D. *Autonome feature extraction and selection*

The retina-like spatial-temporal feature channels are further analyzed to extract low-level features. These binary maps describe the density of edges, irregularity, rough/fine structures, connected structures etc. of the input. The image around the most salient point is processed in detail. Local features are extracted, based on the assumption that the black patches are objects. These objects as entities are collected in a list and their features such as area or eccentricity are computed. Descriptive statistics is used to aggregate the same feature of the different objects such as min or mean.

The number of the features that can be extracted is enormous. We have to find those attributes that are informative enough for proper object categorisation, whereas the number of them is still treatable. We have chosen the Sequential Floating Forward Selection (SFFS) algorithm [15], and the Fisher-quotient as an accuracy function. We have picked this algorithm because in practical adaptations this proved to be the best.

E. *Spatio-temporal event library*

The Event Library contains descriptions of events in the expected scenarios; see Table I. Parallel scenarios are activated by salient features extracted from the scene. If a scenario is active it has an influence on the attention direction. The scenarios are weighted by a priori information and by the identified events, and the more weight a scenario is assigned the bigger the influence it will have on decisions and attention direction.

F. *Multimodal classification with semantic embedding*

The classification task can be greatly enhanced by using semantic embedding. This is the way formally and systematically evaluating the sensory context. These can be location based autonomous tasks, as listed in Table I, or restricted set of objects for example in recognising the number on a public transport vehicle, Figure 4. In addition to the visual input we plan to use auditory clues as well e.g. the noise of the arriving bus or tram, the rustle of the escalator.

There are several classifiers that could have been used. We have applied an adaptive resonance theory (ART) based module, capable of learning on pre-selected training image flows [16]. The ART network has its inspiring roots in neurobiological modeling and has a mathematical back-ground. A further advantage is that a modified version of ART can be implemented on existing CNN-UM architecture.

Acknowledgment

The support of the Hungarian Academy of Sciences, the P. Pázmány Catholic University, the Office of Naval Research as well as the Szentágothai Knowledge Center are kindly acknowledged.

References

[1] L. O. Chua, T. Roska, Cellular Neural Networks and Visual Computing, Cambridge University Press, Cambridge, UK, 2002.

[2] T. Roska, "Computational and Computer Complexity of Analogic Cellular Wave Computers", *Journal of Circuits, Systems, and Computers*, Vol. 12, No. 4, pp. 539-562, 2003

[3] F. S. Werblin, T. Roska and L. O. Chua, "The analogic cellular neural network as a bionic eye," *Intl. J. of Circuit Theory and Applications*; Vol. 23, pp. 541-569, 1995

[4] B. Roska and F. S. Werblin, "Vertical interactions across ten parallel, stacked representations in the mammalian retina," *Nature*, Vol. 410, pp. 583-587, 2001.

[5] A Zarándy, Cs. Rekeczky, P. Földesy, I. Szatmári, "The new framework of applications – The Aladdin system," *J. Circuits Systems Computers* Vol. 12, pp. 769-782, 2003.

[6] Cs. Rekeczky, I. Szatmári, D. Bálya, G. Tímár, and Á. Zarándy, "Cellular Multiadaptive Analogic Architecture: a Computational Framework for UAV Applications," *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol. 51, pp.864-884, 2004

[7] D. Bálya, B. Roska, T. Roska, F. S. Werblin, "A CNN Framework for Modeling Parallel Processing in a Mammalian Retina," *Int'l Journal on Circuit Theory and Applications*, Vol. 30, pp. 363-393, 2002

[8] M. Csapodi and T. Roska, “Dynamic analogic CNN algorithms for a complex recognition task - a first step towards a bionic eyeglass,” *Int. Journal ofCTA*, Vol. .24, No.1, pp.127-144, 1996

[9] V. Gál, T. Roska, “Collision Prediction via the CNN Universal Machine Int.", *Workshop on Cellular Neural Networks and Their Applications* (CNNA 2000), Catania, Italy, pp 105-110.

[10] Á. Zarándy, F. Werblin, T. Roska and L. O. Chua, "Novel Types of Analogic CNN Algorithms for Recognizing Bank-notes," *Proceedings of IEEE Int. Workshop on Cellular Neural Networks and Their Applications*, pp. 273-278, 1994

[11] L. Itti, Modeling Primate Visual Attention, in**:** *Computational Neuroscience: A Comprehensive Approach,* (J. Feng Ed.), pp. 635-655, Boca Raton: CRC Press, 2003.

[12] K. Karacs and T. Roska, “Holistic Feature Extraction from Handwritten Words on Wave Computers”, *Proc. IEEE Int’l Workshop on Cellular Neural Networks and their Applications* (CNNA 2004), pp. 364-369, Budapest 2004

[13] R. Wagner, Á. Zarándy and T. Roska: “Adaptive Perception with Locally-Adaptable Sensor Array”, *IEEE Transactions on Circuits and Systems I, : Regular Papers*, Vol. 51, No.5, pp. 1014-1023, 2004

[14] J. E. Dowling, The Retina: An Approachable Part of the Brain, The Belknap Press of Harvard University Press, Cambridge, 1987.

[15] P. Pudil, F. J. Ferri, J. Novovicova, and J. Kittler, “Floating search methods for feature selection with nonmonotonic criterion functions,” *Proc. Inter. Conf. on Pattern Recognition*, 1994, vol. 1, pp. 279–283.

[16] G. Carpenter and S. Grossberg “A massively parallel architecture for a selforganizing neural pattern recognition machine,” *Computer Vision, Graphics, and Image Processing*, Vol. 37, 1987, pp 54-115.

[17] H.D. Cheng*, X.H. Jiang, Y. Sun, Jingli Wang “Color image segmentation: advances and prospects,” *Pattern Recognition* Vol. 34, (2001), pp 2259-2281

[18] “Web pages about the color space transformations:

”color conversion algorithms”, url = http://jfet.org/color-conversions.html

”color spaces”, url = http://www.couleur.org/index.php?page=transformations.”

[19] Web page about k-means clustering e.g.:

url = http://fconyx.ncifcrf.gov/ lukeb/kmeans.html

[20] T. Roska, D. Bálya, A. Lázár, K. Karacs and R. Wagner, “System aspects of a bionic eyeglass,” in *Proc. of the 2006 IEEE International Symposium on Circuits and Systems (ISCAS 2006)*, Island of Kos, Greece, May 21–24, 2006.

[21] K. Karacs and T. Roska, “Route number recognition via the Bionic Eyeglass,” in *Proc. of 10th IEEE International Workshop on Cellular Neural Networks and their Applications*, Istanbul, Turkey, Aug 28–30, 2006.

[22] E. R. Kandel and J. H. Schwartz, “Principles of Neuroscience” Elsevier, New York, 1985