

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/271834159>

Hand Gesture Based Remote Control System Using Infrared Sensors and a Camera

Article in IEEE Transactions on Consumer Electronics · November 2014

DOI: 10.1109/TCE.2014.7027342

CITATIONS

33

READS

1,740

2 authors:



Fatih Erden

North Carolina State University

50 PUBLICATIONS 295 CITATIONS

SEE PROFILE



A. Enis Cetin

Bilkent University

297 PUBLICATIONS 4,743 CITATIONS

SEE PROFILE

Hand Gesture Based Remote Control System Using Infrared Sensors and a Camera

Fatih Erden and A. Enis Çetin, *Fellow*, IEEE

Abstract — *In this paper, a multimodal hand gesture detection and recognition system using differential Pyro-electric Infrared (PIR) sensors and a regular camera is described. Any movement within the viewing range of the differential PIR sensors are first detected by the sensors and then checked if it is due to a hand gesture or not by video analysis. If the movement is due to a hand, one-dimensional continuous-time signals extracted from the PIR sensors are used to classify/recognize the hand movements in real-time. Classification of different hand gestures by using the differential PIR sensors is carried out by a new winner-take-all (WTA) hash based recognition method. Jaccard distance is used to compare the WTA hash codes extracted from 1-D differential infrared sensor signals. It is experimentally shown that the multimodal system achieves higher recognition rates than the system based on only the on/off decisions of the analog circuitry of the PIR sensors¹.*

Index Terms — **Hand gesture recognition, infrared sensors, sensor fusion, wavelet transform, winner-take-all (WTA) hash method.**

I. INTRODUCTION

Studies on hand and face gesture recognition started in the early 1990s and continues ever since [1]. These studies have many potential applications in the areas of human-computer interaction (HCI), virtual reality (VR), remote control (RC) and industry [2]. Remote and contact-free control of electrical appliances has become a desirable feature in consumer and industrial electronics. Contact-free user interfaces are also useful to reduce the hygiene risks in public places.

This paper focuses on hand gesture recognition using infrared (IR) sensors. Hand gesture recognition systems may be classified into three groups: systems using hand-held pointing devices, systems using wearable sensors and systems doing two dimensional (2-D) image analysis [3]. Generally, it may not be practical to utilize systems requiring gloves or similar equipment for gesture recognition [4]-[6]. It is desirable to have natural hand-gestures to control appliances and TV sets.

¹ Fatih Erden is with the Department of Electrical and Electronics Engineering, Hacettepe University (Beytepe Campus), Ankara 06800 TURKEY (e-mail: erden@ee.bilkent.edu.tr).

A. Enis Çetin is with the Department of Electrical and Electronics Engineering, Bilkent University, Ankara 06800 TURKEY (e-mail: enis@bilkent.edu.tr).

Video based hand gesture recognition is an active research area [7], [8]. Although these approaches provide flexibility in terms of recognition distance and accuracy, they may make false detections due to the illumination changes and reflection. As pointed out by Wachs *et al.* [9], there are only a few practical vision based hand gesture recognition systems.

Current Pyro-electric Infrared (PIR) hand gesture recognition systems are all based on the on/off decisions of the analog circuitry of the PIR sensor [10]-[12]. Wojtczuk *et al.* [13] use a 4x4 PIR sensor array to obtain a switching pattern of PIR sensors to increase the recognition rates. The 4x4 PIR sensor array has the capability to recognize a number of hand gestures but it cannot distinguish face and body gestures from hand gestures, because PIR sensors respond to all hot bodies in their viewing range. Besides an additional camera used for hand detection, the novelty of the multimodal system presented here is the use of continuous-time real-valued signals that PIR sensors produce during an action. As a result it achieves higher recognition rates than the 4x4 PIR sensor array [13] due to the novel use of the analog PIR sensor signals.

The proposed multimodal system consists of a differential PIR sensor array and an ordinary camera for hand gesture recognition. Data obtained from the PIR sensor array and the camera are transferred to a computer and then processed together in real-time. Once any kind of motion is detected by one of the PIR sensors, the motion is checked whether it is due to a hand or not by the camera. If the source of the motion is a hand gesture, data received from the three different PIR sensors are evaluated at the same time to classify the hand gestures e.g., right-to-left/left-to-right, upward/downward hand motions etc. Classification of the hand gestures by the PIR sensor array is carried out by a new winner-take-all (WTA) hash based method. This multimodal solution to the hand gesture detection and recognition problem is a good alternative to the existing methods because of its accuracy, low cost and low power consumption.

This paper extends an earlier study [14] which describes a hand gesture based remote control system using two infrared sensors and camera. The earlier study presents the results produced by the camera-only and the multi-modal systems and reports the improvements made to recognition accuracy of the hand gestures by the multi-modal system. In this paper, the camera is just used for the detection of a hand and the hand gestures are classified using only the sensor array which in this case consists of three differential PIR sensors. By

B. Processing Sensor Output Signals and Decision Mechanism

In this approach, wavelet based signal processing methods are used to extract features from sensor signals. Wavelet domain analysis provides robustness to variations in the sensor signal caused by temperature changes in the environment. In order to keep the computational cost of the detection low, Lagrange filters are used to compute wavelet coefficients.

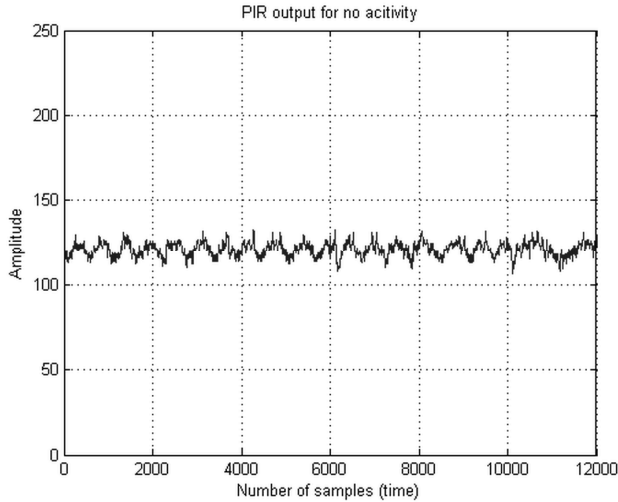


Fig. 2. A typical PIR sensor output signal when there is no activity within its viewing range (sampled at 100 Hz)

Let $x[n]$ be a sampled version of the signal produced by one of the PIR sensors. Wavelet coefficients are obtained after a single stage sub-band decomposition corresponds to [25 Hz, 50 Hz] frequency band information of the original sensor output signal $x[n]$ because the sampling rate is 100 Hz. In this single stage sub-band decomposition, output signal is filtered with an integer arithmetic high-pass filter corresponding to Lagrange wavelets [15] followed by decimation by 2. The transfer function of the high-pass filter- $H(z)$ is given by:

$$H(z) = \frac{1}{2} - \frac{1}{4}(z^{-1} + z^1). \tag{1}$$

The wavelet signals obtained from the two members of the PIR sensor array due to a right-to-left hand motion are shown in Fig. 3. In the hand gesture recognition system, viewing range of the PIR sensors is directionally arranged so that left-to-right, right-to-left, upward, downward, clockwise and counter-clockwise hand movements are determined more accurately. As shown in Fig. 3, the PIR on the right responds earlier than the PIR on the left to a hand motion from right to left. The hand enters the viewing range of the right-most PIR first due to the directional arrangement of the PIR sensors.

Sampled signals from each differential PIR sensor are divided into time windows of length 200 samples covering a 2 seconds period and wavelet coefficient sequences of length 100 corresponding to each window are computed. The demonstrator PIR sensor array system consists of three PIR sensors each of which is one corner of a triangle. Let $w_{r,n}[k]$, $w_{l,n}[k]$ and $w_{u,n}[k]$ $k=(1,\dots,100)$ represent the wavelet coefficient sequences

corresponding to n -th data windows of the PIR sensor signals in the right side, left side and the one on top, respectively. Vectors $w_{r,n}$, $w_{l,n}$ and $w_{u,n}$ are concatenated and a vector $w_{s,n}$, of length 300 for the n -th window is formed. Afterwards this feature vector, which represents all members of the PIR sensor array at the same time, is transformed into binary codes using the WTA hash method. WTA hashing provides a way to convert arbitrary feature vectors into compact binary codes. These codes are resilient to small perturbations in the feature vector. They preserve the rank correlation and can be easily calculated [16]. Computing of a WTA code C_X , corresponding to a feature vector $w_{s,n}$ is explained in Algorithm I.

**ALGORITHM I
WTA HASH ALGORITHM**

- a. Generate random permutation matrices- θ_i s, each of size $M \times M$, $i=1,2,3,\dots,h$.
(M : the length of the vector $w_{s,n}$)
- b. For $i=0:h-1$
 - * $W_i = \theta_i w_{s,n}$
 - * select first K items of W_i
 - * find the index of the maximum item
 - * convert to binary code- c_{x_i}
- c. $C_X = [c_{x_0}, c_{x_1}, \dots, c_{x_{h-1}}]$

Random permutation matrix- θ_i has exactly one entry 1 in each row and each column and 0s elsewhere. θ_i s are generated once at the beginning of the algorithm and then used to permute the wavelet coefficient sequences corresponding to each data window.

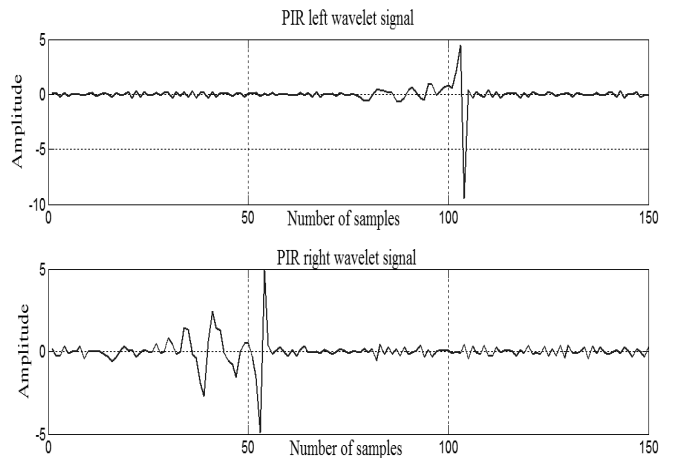


Fig. 3. Wavelet transform of the two PIR sensors signals due to a right-to-left hand motion.

In the classification process of 1-D signals received from the PIR sensors, the method described by Dean *et al.* [17], which calculates a WTA code based Hamming distance metric to find the similarity of images, is used. Hamming distance value depends on the length of the code sequence. Instead, Jaccard distance metric which takes values in the interval [0, 1] independent of the length

of the code sequence is used. The class affiliation of each data window is determined by computing the Jaccard distance between the binary code of the data window and the representative member of each action class. While calculating the Jaccard distance between any C_1 and C_2 binary code sequences of the same length, the contingency table is formed first (Table I).

Parameters p , q , r and s in Table I represent the number of cases $C_{1,i}$ and $C_{2,i}$ are both 1, $C_{1,i}$ is 0 and $C_{2,i}$ is 1, $C_{1,i}$ is 1 and $C_{2,i}$ is 0 and $C_{1,i}$ and $C_{2,i}$ are both 0, respectively. Depending upon these values, the Jaccard distance between C_1 and C_2 sequences is calculated as follows:

$$d_{C_1, C_2} = \frac{q+r}{p+q+r} \quad (2)$$

TABLE I
CONTINGENCY TABLE

		C_1	
		1	0
C_2	1	p	q
	0	r	s

Jaccard distances between the WTA code of $w_{s,n}$ and the codes determined during training phase for each model are calculated and the model yielding the smallest distance is reported as the result of the analysis for the n -th data window.

III. HAND DETECTION USING VIDEO ANALYSIS

Motion detected by the PIR sensors may be due to the movement of a head or whole body of the viewer. Therefore, identification of hand is critical to reduce false alarms. Whenever the analog decision circuitry of the PIR sensor detects motion in the viewing range of one of the PIR sensors, video analysis starts. Standard video analysis methods are used to detect hand within the range of the regular camera. Since the video analysis is controlled by PIR sensors, the resulting system is a low-power consuming system.

First, the presence of a raised hand in the viewing range of the camera is detected. If it is a hand it is possible to recognize the following hand gestures: (i) right-to-left, (ii) left-to-right, (iii) upward, (iv) downward, (v) clockwise, (vi) counter-clockwise hand motions, (vii) open, (viii) closed hand decisions and (ix) counting the number of open fingers. This is useful to change the channels in TV sets and/or set top boxes. One can go to the next channel or the previous one and raise or lower the sound volume by waving his or her hand.

A. Skin Detection Using Classical Color Spaces

Skin colored regions are first detected and then convex hull-defect analysis is performed to recognize the above hand gestures. HSV and YCbCr color spaces are, in general, used for skin detection [18]. In this article YCbCr color space is used, because it is computationally efficient to obtain the YCbCr components from the RGB video color data.

After having the appropriate skin mask, morphological operations, dilation and erosion, are applied to have a cleaner

mask. In dilation phase bright areas within the image are grown by scanning a circle shaped kernel over the image. During the erosion phase the same circle shaped kernel is used. As a result bright areas of the image become thinner after the erosion procedure making the fingers more visible.

B. Convex Hull-Defect Analysis

In this stage, contours extracted from the binary hand mask are processed using the convex hull-defect analysis. After calculating the area of the extracted contours, the largest area is found and assumed to be a 'hand'. By lifting his or her hand closer to the camera than his or her face, the user clearly makes his or her intention to control the system as illustrated in Fig. 4.

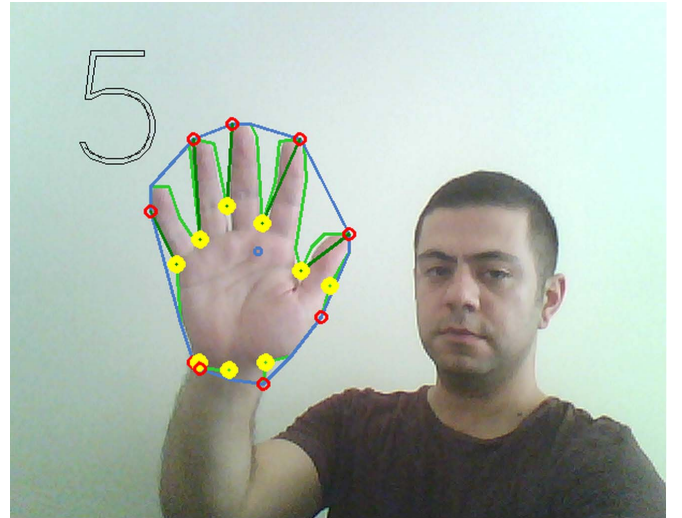


Fig. 4. Hand image and the corresponding convex hull-defect and contour lines.

A pre-defined hand gesture further reduces false alarms. For example the user may open his or her fingers as shown in Fig. 4. This is an unnatural hand gesture therefore it can be used as a remote control signal to the appliance. If the system recognizes more than three fingers, it starts the PIR signal analysis.

Convex defect analysis algorithm proposed by Graham and Yao [19] provides three important parameters: start point, end point and depth information. These parameters represent respectively, the point where defect starts, where it ends and the longest distance between these start and end points. The points and the lines are shown on a hand image in Fig 4. The number of depth lines above a pre-defined length gives the number of open fingers. If the number of open fingers counted is zero, the hand is closed, and if it is one or greater than one, the hand is considered open.

IV. EXPERIMENTS AND RESULTS

The multimodal demonstrator system consisting of three differential PIR sensors and a camera is shown in Fig. 5. It can detect hand and recognize left-to-right, right-to-left, clockwise and counter-clockwise hand gestures. It is also possible to

detect upward/downward motions as well by adding one more PIR sensor aligned with the top PIR sensor into the system.

As pointed in Section III, whenever the user wants to interact with an electrical appliance, he or she raises his or her hand with open fingers in front of the camera. The user can end the controlling action when he or she makes a fist. After detection of a fist by the camera, the multimodal system goes to standby mode and if the user wants to give more commands, he activates the system again by showing his or her hands with open fingers in front of the camera.

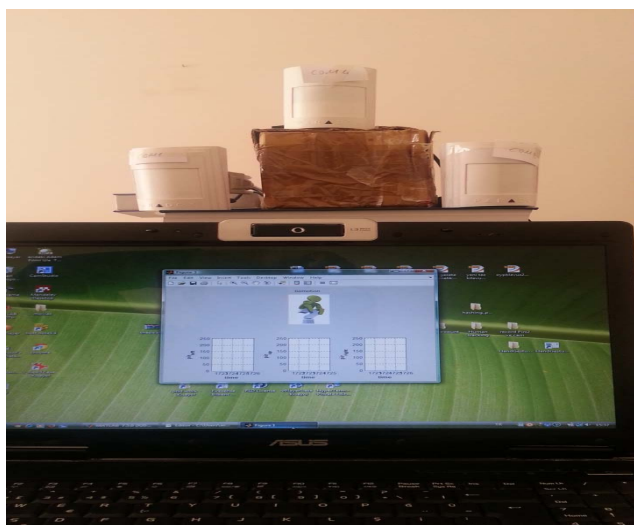


Fig. 5. Setup of the hand gesture detection and recognition system (3 differential PIR sensors and a laptop's camera)

The system can detect and recognize hand gestures up to 1.5 meters. It is assumed that the distance between the user and the TV set is about 1.5 to 2 meters. It is also possible to increase the range using PIR sensors with more directional selectivity and range. There is a Fresnel lens in front of most PIR sensors. The quality of the lens improves the range of the sensor.

TABLE II
CLASSIFICATION RESULTS FOR 312 LEFT-TO-RIGHT/RIGHT-TO-LEFT, 298 UPWARD/DOWNWARD AND 282 CLOCKWISE/COUNTER-CLOCKWISE HAND MOTIONS BELONGING TO SIX DIFFERENT USERS.

Direction	Number of tests	False detections	Accuracy (%)
left/right	312	4	98.7
up/down	298	7	97.6
circular	282	16	94.3

Classification results for 312 left-to-right/right-to-left, 298 upward/downward and 282 clockwise/counter-clockwise hand motions belonging to six different users sitting in front of the multimodal system at a distance of 2 meters are summarized in Table II. The results in the second row of Table II are obtained by rotating the PIR sensor array system by 90 degrees. The accuracy reported by Wojtczuk *et al.* [13] for upward/downward, left-to-right/right-to-left cases is 92.6%. On the other hand the overall accuracy of the proposed multimodal system for these hand motions is 98.2%. The

multimodal system is able to recognize circular motions, i.e. clockwise and counter-clockwise hand motions, as well with a success rate of 94.3%.

V. CONCLUSION

In this article a multimodal hand gesture detection and recognition system is presented. Since both infrared and visible range information is used, the proposed system is more accurate than IR-only and less power consuming than camera-only systems.

A novel WTA code based sensor fusion algorithm is also presented for 1-D PIR sensor signal processing. The algorithm fuses the data coming from the different PIR sensors in an automatic manner to determine left-to-right, right-to-left, upward, downward, clockwise and counter-clockwise motions. A Jaccard distance based metric is used to classify the hash codes of feature vectors extracted from sensor signals.

REFERENCES

- [1] J. M. Rehg and T. Kanade, "Digiteyes: Vision-based hand tracking for human-computer interaction," in *Proc. IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, Texas, USA, pp. 16-22, Nov. 1994.
- [2] W. T. Freeman and C. D. Weissman, "Hand gesture machine control system," *Computer Integrated Manufacturing Systems*, vol. 10, no. 2, pp.175-175, May 1997.
- [3] M. Ishikawa and H. Matsumura, "Recognition of a hand-gesture based on self-organization using a DataGlove," in *Proc. of the 6th International Conference on Neural Information Processing*, vol. 2, pp. 739-745, Perth, Australia, Nov. 1999.
- [4] Y. Han, "A low-cost visual motion data glove as an input device to interpret human hand gestures," *IEEE Trans. Consumer Electron.*, vol. 56, no.2, pp. 501-509, May 2010.
- [5] O. Z. Ozer, O. Ozun, C. O. Tuzel, V. Atalay, and A. E. Cetin, "Vision-based single-stroke character recognition for wearable computing," *IEEE Intell. Syst.*, vol. 16, no. 3, pp. 33-37, Jan./Feb. 2001.
- [6] K. Gill, S. H. Yang, F. Yao, and X. Lu, "A zigbee-based home automation system," *IEEE Trans. Consumer Electron.*, vol. 55, no.2, pp. 422-430, May 2009.
- [7] A. Erdem, E. Erdem, Y. Yardimci, V. Atalay, and A. E. Çetin, "Computer vision based mouse," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Florida, USA, vol. 4, pp. IV-4178, May 2002.
- [8] D. Lee and Y. Park, "Vision-based remote control system by motion detection and open finger counting," *IEEE Trans. Consumer Electron.*, vol. 55, no.4, pp. 2308-2313, Nov. 2009.
- [9] J. P. Wachs, M. Kölsch, H. Stern, and Y. Edan, "Vision-based hand-gesture applications," *Communications of the ACM*, vol. 54, no. 2, pp. 60-71, Jan. 2011.
- [10] R. Samson and T. R. Winings, "Automated dispenser for disinfectant with proximity sensor," U.S. Patent 5,695,091, Dec. 9, 1997.
- [11] F. Erden et al., "Wavelet based flickering flame detector using differential PIR sensors," *Fire Safety Journal*, vol. 53, pp. 13-18, Oct. 2012.
- [12] Y. W. Bai, Z. L. Xie, and Z. H. Li, "Design and implementation of a home embedded surveillance system with ultra-low alert power," *IEEE Trans. Consumer Electron.*, vol. 57, no.1, pp. 153-159, Feb. 2011.
- [13] P. Wojtczuk, A. Armitage, T. D. Binnie, and T. Chamberlain, "Recognition of simple gestures using a PIR sensor array," *Sensors & Transducers*, vol. 14, no. 1, pp. 83-94, Mar. 2012.
- [14] F. Erden, A. S. Bingöl, and A. E. Çetin, "Hand gesture recognition using two differential PIR sensors and a camera," in *Proc. IEEE 22nd Signal Processing and Communications Applications Conference*, Trabzon, Turkey, pp. 349-352, Apr. 2014.
- [15] C. W. Kim, R. Ansari, and A. E. Cetin, "A class of linear-phase regular biorthogonal wavelets," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, San Francisco, USA, vol. 4, pp. 673-676, Mar. 1992.

- [16] J. Yagnik, D. Strelow, D. A. Ross, and R. S. Lin, "The power of comparative reasoning," in *Proc. IEEE International Conference on Computer Vision*, Barcelona, Spain, pp. 2431-2438, Nov. 2011.
- [17] T. Dean, M. A. Ruzon, M. Segal, J. Shlens, S. Vijayanarasimhan, and J. Yagnik, "Fast, accurate detection of 100,000 object classes on a single machine," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, USA, pp. 1814-1821, June 2013.
- [18] R. Khan, A. Hanbury, J. Stöttinger, and A. Bais, "Color based skin classification," *Pattern Recognition Letters*, vol. 33, no.2, pp. 157-163, Jan. 2012.
- [19] R. L. Graham and F. F. Yao, "Finding the convex hull of a simple polygon," *Journal of Algorithms*, vol. 4, no. 4, pp. 324-331, Dec. 1983.

BIOGRAPHIES



F. Erden received his B.Sc and M.S degrees in Electrical and Electronics Engineering from Bilkent University, Ankara, Turkey. Since 2009, he has been a Ph.D. Student in the Department of Electrical and Electronics Engineering at Hacettepe University, Ankara, Turkey. His research interests include sensor signal processing, sensor fusion, multimodal surveillance systems and object tracking.



A. E. Çetin (F'09) studied Electrical Engineering at Middle East Technical University. After getting his B.Sc. degree, he got his M.S.E and Ph.D. degrees in Systems Engineering from the Moore School of Electrical Engineering at the University of Pennsylvania in Philadelphia. Between 1987-1989, he was Assistant Professor of Electrical Engineering at the University of Toronto, Canada. Since then he has been with Bilkent University, Ankara, Turkey. Currently he is a full professor. He spent 1994-1995 academic year at Koç University in Istanbul, and 1996-1997 academic year at the University of Minnesota, Minneapolis, USA as a visiting associate professor. He is involved in Multimedia Understanding Through Semantics and Computational Learning Network of Excellence research MUSCLE-ERCIM, computer vision based wild-fire detection research VBI Lab, biomedical signal and image processing research, MIRACLE project, signal processing research for food safety and quality applications, wavelet theory, inverse problems and used to carry out research related to Turkish Language and Speech Processing. He is currently a member of the SPTM technical committee of the IEEE Signal Processing Society. He founded the Turkish Chapter of the IEEE Signal Processing Society in 1991. He was on the editorial boards of EURASIP Journals, Signal Processing and Journal of Advances in Signal Processing (JASP). Currently, he is the Editor-in-Chief of Signal, Image and Video Processing SIViP, and a member of the editorial boards of IEEE CAS for Video Technology and IEEE Signal Processing Magazine. He holds four US patents.