ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 4, April 2018

Real-Time Elderly People Adl Monitoring

[1] Chidanand.H, [2] Anil Kumar, [3] Anusha A3, [4] Priyanka K
[1][2][3] Department of CSE, RYMEC, Ballari

Abstract: - **This paper proposes a real time human daily activity recognition using a single stationary camera. The system can recognize activities such as raising or waving hand(s), jumping and walking. Actions are described by a feature vector comprising both trajectory information and a set of local motion descriptors. Feature extraction is done using LBP TOP method. Action recognition is achieved via probabilistic search of image feature databases representing previously seen actions. For activity detection and tracking, we use motion detection algorithms such as frame differencing and feature correlation and finally classification is by kNN classifier. Activity recognition can be exploited to great social benefits, especially in real-life, human centric applications such as elder care and healthcare. Recognizing complex activities remains a challenging and active area of research and the nature of human activities poses different challenges.**

Keywords :- **Elderly people Recognition, ADL monitoring, Human activities, LBP TOP.**

I. INTRODUCTION

The availability of a system capable of automatically classifying the physical activity performed by a human subject is extremely attractive for many applications in the field of healthcare monitoring and in developing advanced human-machine interfaces. By the term physical activity, we mean either static postures, such as standing, sitting, lying, or dynamic motions, such as walking, running, stair climbing, cycling, and so forth.

Human activity recognition plays a significant role in human-to-human interaction and interpersonal relations. Because it provides information about the identity of a person, their personality, and psychological state, it is difficult to extract. The human ability to recognize another person's activities is one of the main subjects of study of the scientific areas of computer vision and machine learning. As a result of this research, many applications, including video surveillance systems, human-computer interaction, and robotics for human behavior characterization, require a multiple activity recognition system. Among various classification techniques two main questions arise: "What action?" (i.e., the recognition problem) and "Where in the video?" (i.e., the localization problem). When attempting to recognize human activities, one must determine the kinetic states of a person, so that the computer can efficiently recognize this activity. Most of the work in human activity recognition assumes a figure-centric scene of uncluttered background, where the actor is free to perform an activity. The development of a fully automated human activity recognition system, capable of classifying a person's activities with low error, is a challenging task due to problems, such as background clutter, partial occlusion, changes in scale, viewpoint, lighting and appearance, and frame resolution. In addition, annotating behavioural roles is time consuming and requires knowledge of the specific event. Moreover, intra- and interclass similarities make the problem amply challenging. That is, actions within the same class may be expressed by different people with different body movements, and actions between different classes may be difficult to distinguish as they may be represented by similar information. The way that humans perform an activity depends on their habits, and this makes the problem of identifying the underlying activity quite difficult to determine. Also, the construction of a visual model for learning and analysing human movements in real time with inadequate benchmark datasets for evaluation is challenging tasks.

II LITERATURE SURVEY

1.Volker Kruger [5], categorised and reviewed on the basis of scene, full body and with/without using body parts.

2. Jose M. Chaquet and Enrique J. Carmona reviewed 68 datasets available for human action recognition within the time interval of 2001 to 2012. Teofilo E. de Campos[7] survey focus on the applications of the human action recognition in video survilliance for observing people in real time

Figure: Mobile Receiving Unit

III WORKING NATURE

The block diagram of the project "Human Activity Recognition Using Machine Learning Technique and Alert System for Health Care in Hospitals" is as shown in the Figure.

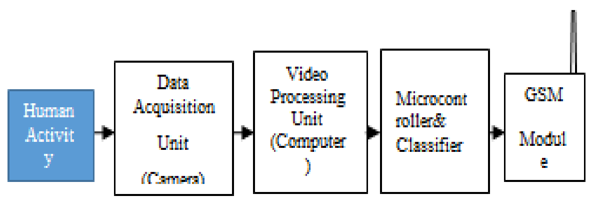


Figure: Block diagram Transmitting Unit

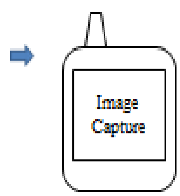


Figure: Block diagram Transmitting Unit

SYSTEM HARDWARE

This chapter deals with the description of overall circuitry of project and details of the hardware component used.

The different hardware components used in the system are as follows

- Data acquisition unit
- Camera
- Data processing unit
- LBP TOP for feature extraction
- kNN classifier for classification
- Controlling unit
  - 89S52
- Serial port interface
- Electrical appliance
- RF Module
- Power supply

DATA ACQUISITION UNIT:

This unit is used to capture the video of the body movement and give it to the image processing unit.

CAMERA:

A simple wired camera is used to capture video and is connected to the image processing system.

DATA PROCESSING UNIT:

It involves feature extraction and classification. Feature extraction is done using LBP TOP method and classification is by kNN classifier.

LOCAL BINARY PATTERN – TOP:

Local binary pattern operator is very simple efficient texture which is used for labeling the individual pixel of an image by thresholding the neighborhood of individual pixel and results are considered in the form of binary numbers.LBP texture operator has been found in variety of applications and it is considered to be a popular approach because of computational simplicity and discriminative power. In real world applications, the important property of local binary pattern is its robustness to monotonic gray-scale changes caused by illumination variations. It is possible to analyze images in real time because of its computational simplicity.

Local Spatiotemporal Descriptors for Visual Information

Local binary patterns operator is known for its classification for different kinds of textures. Binary code is produced for an individual pixel in an image by thresholding the value of neighborhood pixel with the value of center pixel as shown in Figure 3.1(a) and equation (1)[6]

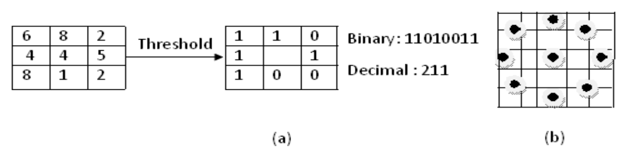


Figure 3.1 (a) Basic LBP operator (b) Circular (8, 2) neighborhood

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p,\; s'(x) = \begin{cases} 1, & x>0 \\ 0, & x<0 \end{cases} \quad \text{.......(1)}$$

Where, gc→ represents gray value of center pixel of local Neighborhood,

g_p→ represents gray values of P equally spaced pixels on a circle of radius R.

By taking into account the differences between the neighborhood values and center pixel values, local binary pattern achieves invariance with respect to gray scale.

A histogram is used to collect the different binary pattern that occurs. In order to include circular neighborhoods with any number of pixels, definition of neighbor can be extended as shown in Figure 3.1(b). So larger scale texture primitives or micro patterns like lines, spots and corners can be collected [1][2]. To shorten the feature vector length of the local binary pattern, uniform patterns are used. If a pattern contains at least two bitwise transitions from 0 to 1 or vice-versa than it are considered

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 4, April 2018

to be uniform. For example 10101000 is not uniform pattern because it has 4-bitwise transition. While considering only uniform local binary pattern, non-uniform patterns are made to store in a single bin during computation of histogram.

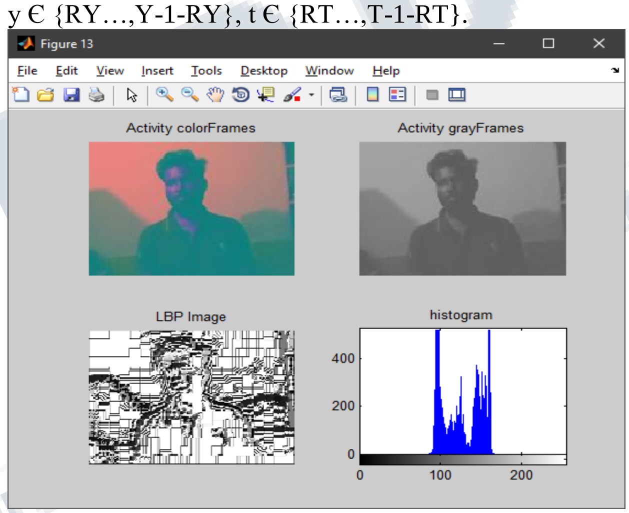
Local texture descriptors have obtained tremendous attention in the analysis of facial image because of their robustness to challenge such as pose and illumination changes. In our approach a temporal texture recognition using local binary patterns was extracted from the three orthogonal planes (LBP-TOP). LBP-TOP method is more efficient than the ordinary LBP. In ordinary LBP we are extracting information or features in two dimensions, where as in LBP-TOP we are extracting information in three dimension i.e., X, Y and T. For LBP-TOP, the radii in spatial and temporal axes X, Y, and T, and the number of neighboring points in the XY, XT, and YT planes can also be different and can be marked as R_X, R_Y and R_T, P_{XY}, P_{XT}, P_{YT}. The LBP-TOP feature is then denoted as LBP-TOP $_{PXY, PXT, PYT, RX, RY, RT}$. If the coordinates of the center pixel $g_{tc,c}$ are (x_c, y_c, t_c), then the coordinates of local neighborhood in XY plane $g_{XY,p}$ are given by (x_c-$R_X\sin(2\pi p/P_{XY})$, y_c+$R_Y\cos(2\pi p/P_{XY})$, t_c), the coordinates of local neighborhood in XT plane $g_{XT,p}$ are given by (x_c-$R_X\sin(2\pi p/P_{XT})$, y_c, t_c –$R_T\cos(2\pi p/P_{XT})$) and the coordinates of local neighborhood in YT plane $g_{YT,p}$(x_c, y_c-$R_Y\cos(2\pi p/P_{YT})$, t_c –$R_T\sin(2\pi p/P_{YT})$). Sometimes, the radii in three axes are the same and so do the number of neighboring points in XY, XT, and YT planes. In that case, we use LBP-TOP$_{P,R}$ for abbreviation where $P=P_{XY}=P_{XT}=P_{YT}$ and $R=R_X=R_Y=R_T$[1][3]. LBP-Top is used to represent the movements of body because of its ability to describe the spatiotemporal signals, robustness to monotonic gray-scale changes caused by illumination variation. Considering the movement of body, the descriptors are obtained by concatenating local binary patterns on three orthogonal planes from the activity sequence: XY, XT, and YT, considering only the co-occurrence statistics in these three directions.

The LBP-TOP histograms in each block volume are computed and concatenated into a single histogram. All features extracted from each block volume are connected to represent the appearance and motion of the body sequence. In this way, we effectively have a description of the body movement on three different levels of locality. The labels (bins) in the histogram contain information from three orthogonal planes, describing appearance and temporal information at the pixel level. The labels are summed over a small block to produce information on a regional level expressing the characteristics of the appearance and motion in specific locations and time segments, and all information from the regional level is concatenated to build a global description of the body motion. Moreover, even though different activities have different length, they are divided into the same number of block volumes, so the lengths of their feature vectors are the same to compare.

A histogram of the mouth movements can be defined as follows [1]:

$$H_{b,c,d,j,i} = \sum_{x,y,t} I\{f_j(x,y,t)=i\},\ i=0,\ldots,n_j-1;\ j=0,1,2 \quad (2)$$

where, n_j→ represents number of different labels produced by the LBP operator in the j[th] plane (j=0: XY, 1:XT, and 2:YT), $f_j(x,y,t)$ expresses the LBP code of central pixel (x,y,t) in the j[th]plane, x Є {RX…,X-1-RX}, y Є {RY…,Y-1-RY}, t Є {RT…,T-1-RT}.



k-Nearest Neighbor Classifier

For recognition of human activity k-Nearest-Neighbour (kNN) [8] classification is used because it is one of the most fundamental and simple classification methods and should be one of the first choices for a classification study when there is little or no prior knowledge about the distribution of the data. K-nearest-neighbour classification was developed from the need to perform discriminant analysis when reliable parametric estimates of probability densities are unknown or difficult to determine.

k-NN assumes that the data is in a feature space. More exactly, the data points are in a metric space. The data can be scalars or possibly even multidimensional vectors. Since the points are in feature space, they have a notion of distance – this need not necessarilybe euclidean distance although it is the one commonly used. Each of the training data consists of a set of vectors and class label associated with each vector. In the simplest case , it

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 4, April 2018

will be either + or – (for positive or negative classes). But KNN , can work equally well with arbitrary number of classes.

Single number "k"decides how many neighbors (where neighbors is defined based on the distance metric) influence the classification. This is usually a odd number if the number of classes is 2. If k=1 , then the algorithm is simply called the nearest neighbor algorithm.

The algorithm has different behavior based on k is as follows:

Case 1 : k = 1 or Nearest Neighbor Rule
Let x be the point to be labeled . Find the point closest to x . Let it be y. Now nearest neighbor rule asks to assign the label of y to x.Consider a point x in the subspace which also has a lot of neighbors. Now let y be the nearest neighbor. If x and y are sufficiently close, then we can assume that probability that x and y belong to same class is fairly same – then by decision theory, x and y have the same class.

Case 2 : k = K or k-Nearest Neighbor Rule
Here we try to find the k nearest neighbour and do a majority voting. Typically k is odd when the number of classes is 2. Let's say k = 5 and there are 3 instances of C1 and 2 instances of C2. In this case, KNN says that new point has to label as C1 as it forms the majority. Follow a similar argument when there are multiple classes.

MICROCONTROLLER UNIT:

The AT89S52 microcontroller is used for programming in order to control the appliances. AT89S52 is a low-power, high-performance CMOS 8-bit microcontroller with 8K bytes of in-system programmable Flash memory. The device is manufactured using Atmel's high-density non-volatile memory technology and is compatible with the industry-standard 80C51 instruction set and pin out. The on-chip Flash allows the program memory to be reprogrammed in-system or by a conventional non-volatile memory programmer. By combining a versatile 8-bit CPU with in-system programmable Flash on a monolithic chip, the Atmel AT89S52 is a powerful microcontroller, which provides a highly flexible and cost-effective solution to many, embedded control applications. The AT89S52 provides the following standard features: 8K bytes of Flash, 256 bytes of RAM, 32 I/O lines, Watchdog timer, two data pointers, three 16-bit timer/counters, a six-vector two-level interrupt architecture, a full duplex serial port, on-chip oscillator, and clock circuitry. In addition, the AT89S52 is designed with static logic for operation down to zero frequency and supports two software selectable power saving modes. The Idle Mode stops the CPU while allowing the RAM, timer/counters, serial port, and interrupt system to continue functioning. The Power-down mode saves the RAM contents but freezes the oscillator, disabling all other chip functions until the next interrupt.

The following are some of the capabilities of 8051 microcontroller.

- Internal ROM and RAM
- I/O ports with programmable pins
- Timers and counters
- Serial data communication

The 8051 architecture consists of these specific features:

- 16 bit PC and data pointer (DPTR)
- 8 bit program status word (PSW)
- 8 bit stack pointer (SP)
- Internal ROM 4K
- Internal RAM of 128 bytes
- 4 register banks, each containing 8 registers

1. 80 bits of general purpose data memory
2. 32 input/output pins arranged as four 8 bit ports: P0-P3
3. Two 16 bit timer/counters: T0-T1
4. Two external and three internal interrupt sources

Oscillator and clock circuits.

SERIAL PORT INTERFACE

The mode communication between pc and microcontroller is serial.

MAX 232

MAX232 IC is a line driver which has two receivers (converts from RS232 to TTL voltage levels) and two drivers (converts from TTL logic to RS232 voltage levels) and is used for interfacing the microcontroller to the computer. This is because the logical'0' and logical'1' in a controller are represented by 0v and 5v respectively while the same is represented by a voltage range of +3v to +12v and -3v to -12v in the serial port of computer. This voltage conversion is done by the line driver so that there is a proper voltage interface between them.

RS232

The RS232 standard interfaces between two different types of equipment, Data Terminal Equipment (DTE) and Data Communication Equipment (DCE). The DTE is microcontroller, which receiver data and DCE is

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 4, April 2018

Max6232, which is responsible for transferring data. A simplex asynchronous data transmission is used; i.e. PC will send data to micro controller through RS232 .
The voltage levels for RS-232 standard is -3v to -25v at logic high level and logic level low of + 3v to +15v and +3v to +25v.

GSM MODULE
GSM is a mobile communication modem; it is stands for global system for mobile communication (GSM). The idea of GSM was developed at Bell Laboratories in 1970. It is widely used mobile communication system in the world. GSM is an open and digital cellular technology used for transmitting mobile voice and data services operates at the 850MHz, 900MHz, 1800MHz and 1900MHz frequency bands.GSM system was developed as a digital system using time division multiple access (TDMA) technique for communication purpose. A GSM digitizes and reduces the data, then sends it down through a channel with two different streams of client data, each in its own particular time slot. The digital system has an ability to carry 64 kbps to 120 Mbps of data rates.

GSM MODULE
Power supply:
In this project 5V and 12V power supply units are designed to power controlling unit and processing unit. The block diagram of power supply is as shown in Figure 3.6 and circuit diagram shown in Figure 3.7

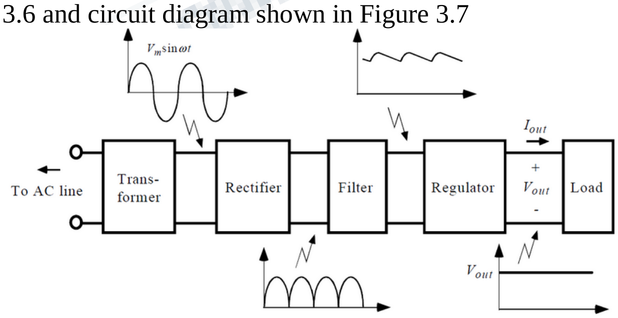


Figure: Block diagram of power supply

SYSTEM DETAILS
This system consists of three stages, as shown in Figure below. The first stage is capturing of human activity movements using webcam. The second stage is extraction of the visual features from the movement sequence. The role of the final stage is to recognize the input activity using a k-NN classifier.

Software Flow for Human Activity Recognition in Matlab

IV SYSTEM SOFTWARE

This chapter has deals with detailed description of the software's which were used in this project. This project is implemented using the following software

- MATLAB
- Keil µvision

MATLAB
MATLAB [5] (matrix laboratory) is a multi-paradigm numerical computing environment and fourth generation programming language. Developed by math works, Matlab allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces and interfacing with programs written in other languages C, C++ java and Forton. Various toolboxes are present in MATLAB. In this project image processing toolbox is used to perform various image processing techniques.

Webcam Support
One can use MATLAB Webcam support to bring live images from any USBVideo Class (UVC) compliant into MATLAB. This includes Webcams that may be built into laptops or other devices, as well as Webcams that plug into your computer via a USB port. Using simple MATLAB functions, can detect connected Webcams, acquire individual snapshots from a Webcam, and optionally set up a loop for acquiring images. The webcam list function allows detecting the connected Webcams. The webcam function creates the Webcam object that is used to acquire images. And the snapshot function returns a single image from the camera.

Table Supported functions of webcam

Function	Purpose
Webcam list	Returns list of Webcams that are connected to the system.
Webcam	Creates webcam object and connects to the single camera on your system. If you have multiple cameras and you use

	the webcam function with no input argument, it creates the object and connects it to the first camera it finds.
Preview	Preview the images from the Webcam. Use name of objects input argument, which is cam in this example: preview(cam)
Snapshot	Acquire a single image from the Webcam. Use name of object as input argument, which is cam in this example. imp = snapshot(cam);

Keil µ(micro) vision :
Keil development tools for the 8051 Microcontroller Architecture support every level of software developer from the professional applications engineer to the student just learning about embedded software development. The Keil 8051 Development Tools are designed to solve the complex problems facing embedded software developers. The Keil µVision Debugger accurately simulates on-chip peripherals of the 8051 device. Simulation helps us to understand hardware configurations and avoids time wasted on setup problems. Additionally, with simulation, write and test applications before target hardware is available.

Flow Chart
The overall operation for human activity detection is explained in the flow chart as shown in Figure

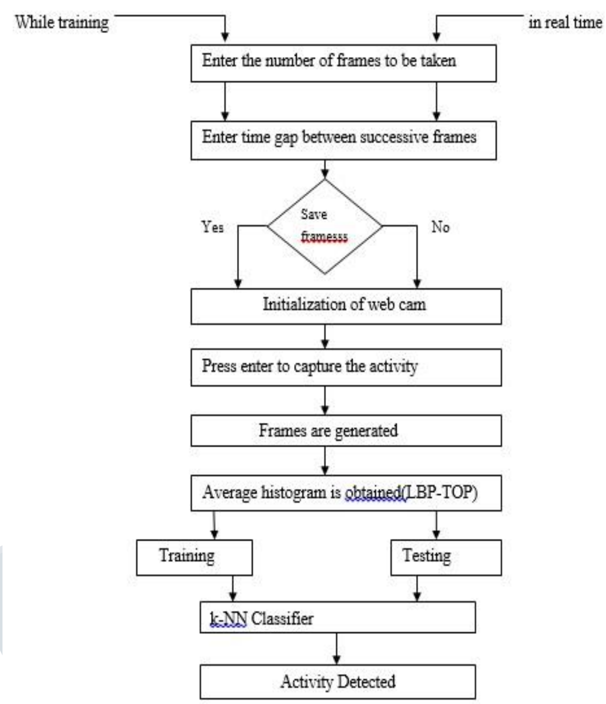


Figure : Flow chart

V TEST AND RESULTS

Initially the system is trained for different activities and in real time the data is obtained and compared with presorted data (while training).

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 4, April 2018

VI ADVANTAGES AND DISADVANTAGES

Advantages:

- It can be used in Hospitals.
- It can be used in Elder care homes.
- It can be used in military applications.

Disadvantages:

- Accuracy is less if the system is trained for more number of activities.

VI. CONCLUSION

The objective of this work is to recognize human activities using a simple webcam instead of costly multiple sensors and complicated algorithms. The overall operation is initialized with image acquisition. It is implemented by capturing video of body movements of the person. Features are extracted from image and stored as histogram. In classification image is compared with histogram from database. The activity detected is displayed and voice is enabled for the same. Hence this system is assistance to elders and patients care takers.

VII. FUTURE UPGRADES

Although the current results are encouraging, present algorithm still has room for improvement. Future plan is to have research to improve the accuracy of recognition of human activity. First, automated methods are needed to find optimal parameter settings during training. Second, recognition rate could be improved using more sophisticated criteria. Third, to improve the recognition rate of user independent experiments. Fourth, faster algorithms and other classifiers will also be investigated.

REFERENCES

[1]Aggarwal, J. K., and Cai, Q. (1999). Human motion analysis: a review. Comput. Vis. Image Understand. 73, 428–440. doi:10.1006/cviu.1998.0744 CrossRef Full Text | Google Scholar

[2]Aggarwal, J. K., and Ryoo, M. S. (2011). Human activity analysis: a review. ACM Comput.Surv. 43, 1–43. doi:10.1145/1922649.1922653 CrossRef Full Text | Google Scholar

[3]Aggarwal, J. K., and Xia, L. (2014). Human activity recognition from 3D data: a review. Pattern Recognit.Lett. 48, 70–80. doi:10.1016/j.patrec.2014.04.011 CrossRef Full Text | Google Scholar

[4]Akata, Z., Perronnin, F., Harchaoui, Z., and Schmid, C. (2013).“Label-embedding for attribute-based classification,” in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Portland, OR), 819–826. Google Scholar

[5]Alahi, A., Ramanathan, V., and Fei-Fei, L. (2014).“Socially-aware large-scale crowd forecasting,” in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Columbus, OH), 2211–2218. Google Scholar

[6]AlZoubi, O., Fossati, D., D’Mello, S. K., and Calvo, R. A. (2013).“Affect detection and classification from the non-stationary physiological data,” in Proc. International Conference on Machine Learning and Applications (Portland, OR), 240–245. Google Scholar

[7]Amer, M. R., and Todorovic, S. (2012). "Sum-product networks for modeling activities with stochastic structure," in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Providence, RI), 1314–1321.
Google Scholar

[8]Amin, S., Andriluka, M., Rohrbach, M., and Schiele, B. (2013)."Multi-view pictorial structures for 3D human pose estimation," in Proc. British Machine Vision Conference (Bristol), 1–12.
Google Scholar

[9]Andriluka, M., Pishchulin, L., Gehler, P. V., and Schiele, B. (2014)."2D human pose estimation: new benchmark and state of the art analysis," in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Columbus, OH), 3686–3693.
Google Scholar

[10]Andriluka, M., and Sigal, L. (2012). "Human context: modeling human-human interactions for monocular 3D pose estimation," in Proc. International Conference on Articulated Motion and Deformable Objects (Mallorca: Springer-Verlag), 260–272.
Google Scholar

[11]Anirudh, R., Turaga, P., Su, J., and Srivastava, A. (2015)."Elastic functional coding of human actions: from vector-fields to latent variables," in Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Boston, MA), 3147–3155.
Google Scholar

[12]Atrey, P. K., Hossain, M. A., El-Saddik, A., and Kankanhalli, M. S. (2010). Multimodal fusion for multimedia analysis: a survey. Multimed. Syst. 16, 345–379. doi:10.1007/s00530-010-0182-0