Towards Classification Based Human Activity Recognition in Video Sequences

Nguyen Thanh Binh[1(✉)], Swati Nigam[2], and Ashish Khare[2]

[1] Faculty of Computer Science and Engineering, Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam
ntbinh@cse.hcmut.edu.vn

[2] Department of Electronics and Communication, University of Allahabad, Allahabad, India
swatinigam.au@gmail.com, ashishkhare@hotmail.com

Abstract. Recognizing human activities is an important component of a context aware system. In this paper, we propose a classification based human activity recognition approach. This approach recognizes different human activities based on a local shape feature descriptor and pattern classifier. We have used a novel local shape feature descriptor which is integration of central moments and local binary patterns. Classifier used is flexible binary support vector machine. Experimental evaluations have been performed on standard Weizmann activity video dataset. Six different activities have been considered for evaluation of the proposed method. Two activities have been selected at a time with binary classifier. These are walk-run, bend-jump, and jack-skip pairs. Experimental results and comparisons with other methods, demonstrate that the proposed method performs well and it is capable of recognizing six different human activities in videos.

Keywords: Human activity recognition · Classification · Feature descriptors

1 Introduction

Successful recognition of human activities enables a wide range of applications. Context aware human activity recognition generally refers to the determination of a human user's activity using camera and other contextual information that is readily available and accessible [1, 2]. The availability and accessibility of this real time contextual information creates new aspects of human activity recognition, as evidenced by Opportunity Activity Recognition Challenge [3] and Nokia Mobile Data Challenge [4]. Recognition of human activities is the final and major step of a complete human behavior analysis system and follows moving object segmentation [5], object recognition [6] and object tracking [7]. Activity recognition is important for many potential applications such as visual surveillance, biomechanical applications, human-computer interaction, clinical purposes and sports analysis, etc.

Generally, human activities can be categorized into four classes [8]. They are (i) human actions: that involve single human, (ii) human-human interactions: that involve two humans, (iii) human-object interactions: that involve human and object,

P.C. Vinh et al. (Eds.): ICCASA 2013, LNICST 128, pp. 209–218, 2014.
DOI: 10.1007/978-3-319-05939-6_21,

and (iv) group activities: that involve multiple humans. But this is a difficult problem due to several challenges. Robustness against environment variations, robustness against actor's movement variations, robustness against various activities and insufficient amount of training videos are few of them [8].

While a number of activity recognition approaches have been explored to deal with above challenges, the proposed method takes advantage of holistic features based local descriptors and pattern classifier. In this paper, we have discussed a simple and effective unimodal approach of human activity recognition. The motivation is that the holistic features based local descriptors emphasize different aspects of activities and are suitable for different types of activity datasets. We have extracted a novel feature descriptor which is integration of moment features upon local binary features. This feature has further been classified using kernel based support vector machine classifier. While sequential classifiers, e.g. HMM [9], are common for learning from sequences, in this paper, we focus on developing a technique that enable using standard non-sequential learning technique for accurate activity recognition. This is motivated by the fact [10] that the non-sequential techniques, such as support vector machine, have good competitiveness and scalability on large-dimensional and continuous-valued activity data. Our experiments are conducted on real-world, publicly available Weizmann activity recognition video dataset [11] and we have considered six common activities and two at a time that are walk-run, bend-jump, and jack-skip pairs. There are several available activity datasets, among which the Weizmann dataset has been widely used to evaluate activity recognition approaches and many results have been reported on it.

The rest of the paper is organized as follows. Section 2 briefly explains the related work in this area, Sect. 3 describes fundamentals of the proposed feature descriptor based method, Sect. 4 explains the proposed method, Sect. 5 elaborates the experimental results and finally Sect. 6 is the conclusion of this paper.

2 Related Work

The main idea of this paper is to use holistic features based local descriptors for classification to perform human activity recognition. Therefore, we review representative papers on these features.

Several features based activity recognition techniques have been developed so far. Variants of SIFT feature, for example SIFT Flow [12] and three dimensional SIFT descriptor [13] have been successfully used for this purpose. The work done in the field of features based activity recognition can be categorized into region based and boundary based approaches. Fourier descriptor [14] is a popular boundary based feature descriptor used in literature. However, region based approaches retrieve information from boundary as well as from internal pixels. Therefore, they are preferred for general shape representation over boundary based approaches. Central moment [15] is another example of region based feature descriptors which has been used in recognition. Moment invariants [16] are well known region based features used for translation, rotation and scale invariant recognition. Flusser and Suk [17] proposed a blur and affine invariant approach for use in pattern recognition applications, e.g. activity recognition. Zernike moments [18] are rotation invariant and Legendre moments [19] are translation and scale invariant techniques for recognition.

Recognition in real scenes is a tough task due to variation in clothes, illumination conditions and poses. Dalal and Triggs [20] presented a method based on Histogram of Oriented Gradients (HOG) for handling these problems. This approach combines histogram of oriented gradients (HOG) and linear support vector machine (SVM) and provides better results. The HOG descriptors were computed and concatenated over detection window of size 16 × 16 pixels blocks. However, since the detection window slides over entire image, therefore this method is computationally expensive. Method to speed up this algorithm has been proposed in [21]. The variants of HOG include circular HOG [22] and hybrid HOG [23].

In recent years, interest is increasing to obtain local patterns of an image for better recognition. Local binary patterns (LBP) with adaptive threshold have provided excellent results for different vision applications [24]. In comparison to the gradient based features like HOG, LBP is more accurate, sparse and provides simple calculation. Moreover, LBP features can work in local color configuration and provide illumination invariant recognition. Motivated by the success of LBP features, [25] presented several variants of the original local binary pattern (LBP). To explore the discriminative ability of LBP, a window based descriptor has been proposed [26] for robust multi viewpoint recognition in different pose and under realistic environment. Center symmetric LBP in wavelet domain [27] has been computed for rapid recognition. An improved and hybrid strategy of HOG and LBP was presented in [28] for efficient and accurate recognition in real scenes.

Hence, it is clear that feature descriptors play a vital role in classification based activity recognition can be effectively used for accurate recognition of human activities.

3 Preliminaries

The purpose of this section is to provide a brief description of the Local Binary Patterns (LBP) and Central Moments (moments) used in the proposed work and to show their effectiveness towards human activity recognition.

3.1 Local Binary Patterns (LBP)

For computation of LBP, let us take an image $f(x, y)$ and g_c represent gray value of a pixel at location (x, y), i.e. $g_c = f(x_c, y_c)$. Also, let g_p represent gray level of any sample point in an equispaced circular neighborhood where P is the number of sample points within radius R. Then the general local binary pattern operator $LBP_{P,R}$ can be obtained as follows

$$LBP_{P,R}(x_c, y_c) = \sum_{P=0}^{P-1} s(g_p - g_c)2^P \tag{1}$$

where $s(u)$ is defined as

$$s(u) = \left\{ \begin{array}{l} 1, u \geq 0 \\ 0, u < 0 \end{array} \right\} \tag{2}$$

3.2 Central Moments (Moments)

The geometric moment m_{pq} of order $p + q$ of an object in an image $f(x, y)$ can be defined as follows

$$m_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} x^p y^q f(x, y) \tag{3}$$

In case of central moments

$$\mu_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (x - x_c)^p (y - y_c)^q f(x, y) \tag{4}$$

$$\text{where } x_c = \frac{m_{10}}{m_{00}} \text{ and } y_c = \frac{m_{01}}{m_{00}} \tag{5}$$

4 The Proposed Method

The steps of the proposed method are explained in this section.

Step 1: Read frames of a video sequence.
Step 2: Scale resize each frame.
Step 3: Convert it from RGB to gray level color space.
Step 4: Compute LBP according to following equation

$$LBP_{P,R}(x, y) = \sum_{P=0}^{P-1} s(g_p - g_c) 2^p, \; s(u) = \left\{ \begin{array}{l} 1, u \geq 0 \\ 0, u < 0 \end{array} \right\} \tag{6}$$

where (x_c, y_c) is the location of center pixel in the region R, g_c is the gray value of center pixel, g_p is the gray value of the neighborhood pixels and P can take values from 0 to $N - 1$. Here N = 8.

Step 5: Compute moment-LBP according to following equation

$$MLBP_{P,R}(x, y) = \sum_{c=0}^{M-1} \sum_{c=0}^{N-1} (x - x_c)^u (y - y_c)^v LBP_{P,R}(x, y) \tag{7}$$

where $u + v = 2$.

Step 6: Classify moment-LBP features. We have computed these features in each and every frame of activity video sequence. The classification has been performed using 'rbf' kernel based binary SVM classifier. This is a two class classifier

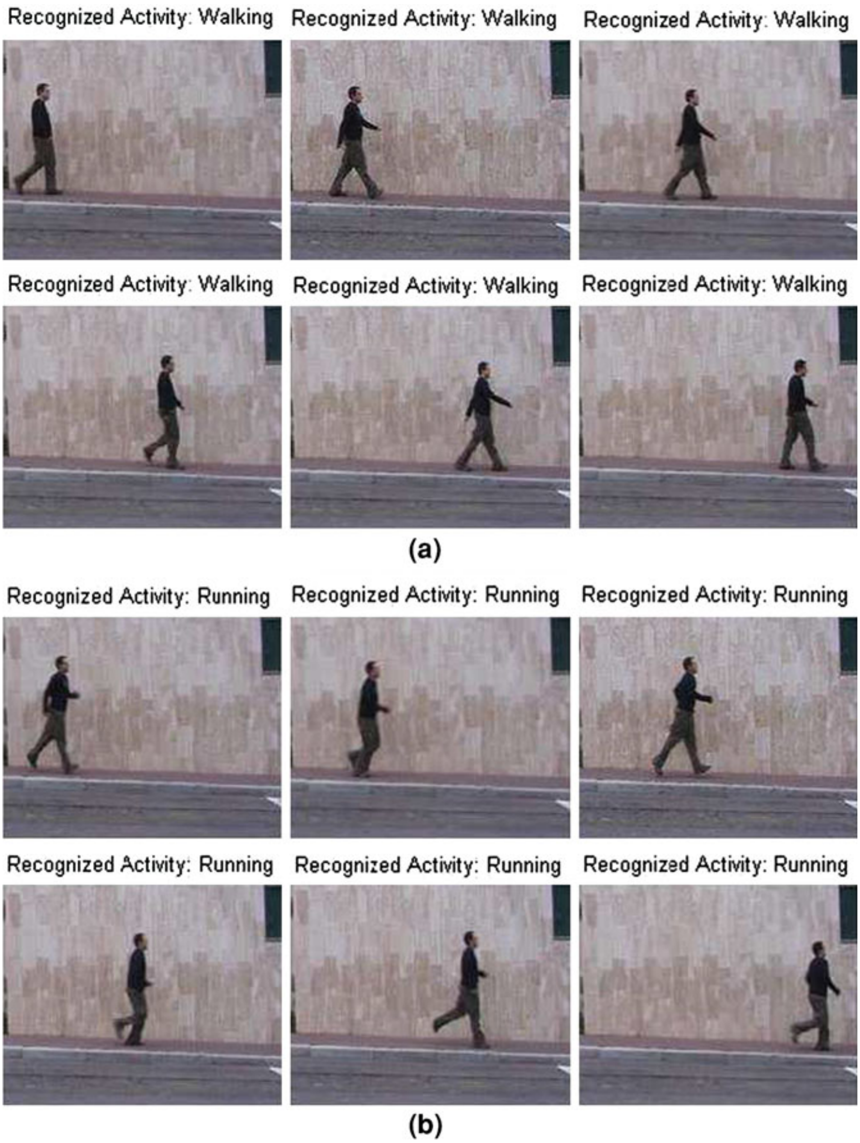


Fig. 1. Results of the proposed method for activities (a) Walk (b) Run (c) Bend (d) Jump (e) Jack and (f) Skip

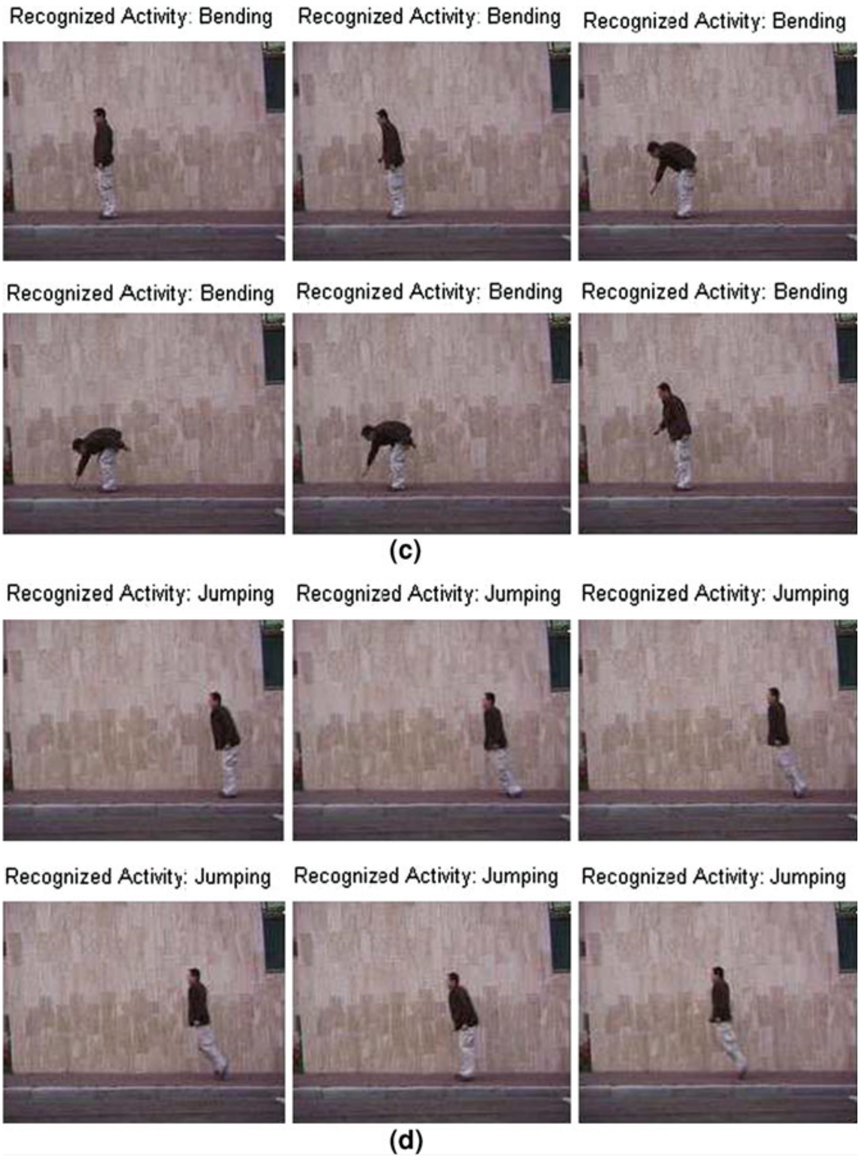
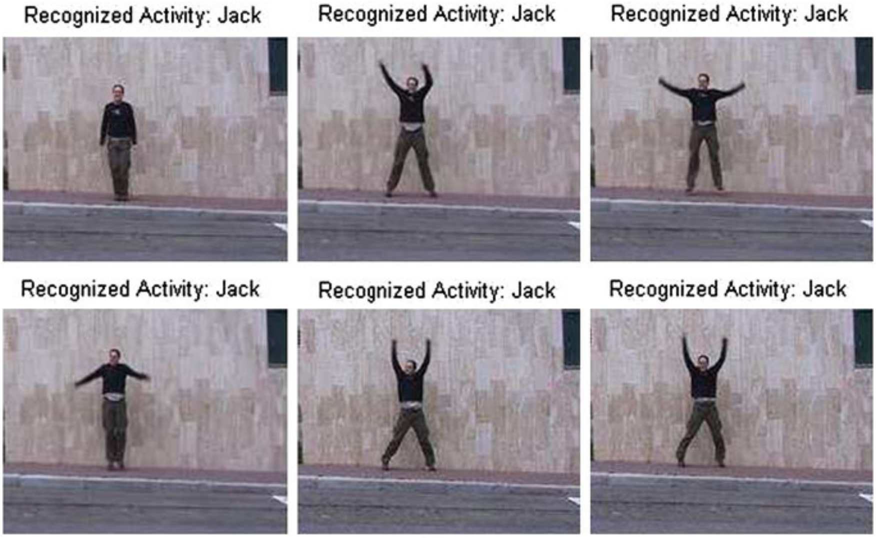
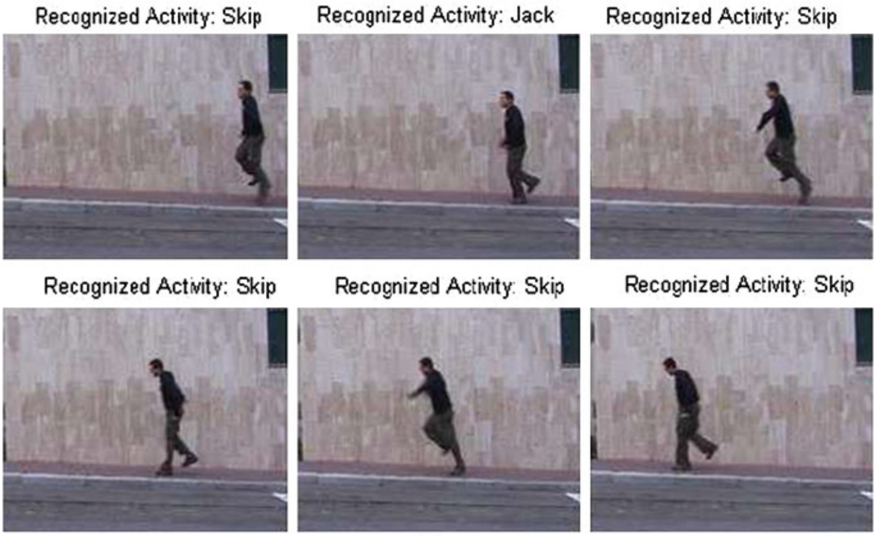


Fig. 1. (continued)



(e)



(f)

Fig. 1. (continued)

and efficiently classifies features for two activities at a time. Pair of activities have been selected for this purpose. These pairs are walk-run, bend-jump and jack-skip. First, we have selected walk-run pair of activities. Features for these activities have been computed separately. Now these features have been classified with binary SVM classifier which is an established classifier for this purpose. Other pair of activities which we have taken are bend-jump and jack-skip from Weizmann dataset.

Table 1. Average accuracy of different methods for activities walk-run pair

Method	Average accuracy (%)
Moment [15]	50.0
LBP [24]	50.0
CSLBP [27]	51.9
HOG [20]	50.0
Proposed	**72.6**

Table 2. Average accuracy of different methods for activities bend-jump pair

Method	Average accuracy (%)
Moment [15]	48.7
LBP [24]	49.1
CSLBP [27]	43.5
HOG [20]	49.1
Proposed	**56.8**

Table 3. Average accuracy of different methods for activities jack-skip pair

Method	Average accuracy (%)
Moment [15]	48.9
LBP [24]	49.5
CSLBP [27]	74.1
HOG [20]	49.5
Proposed	**77.3**

5 Experimental Results

In this section, we have performed experimental evaluations of the proposed method. The proposed method has been performed over Weizmann activity dataset and we have considered 6 different activities, two at a time, that are walk-run, bend-jump, and jack-skip pair. Also, the proposed method has been compared with Moment [15], Local Binary Patterns (LBP) [24], Central Symmetric Local Binary Patterns (CSLBP) [27] and Histogram of Oriented Gradients (HOG) [20] feature descriptor based methods. Figure 1 visually shows results of the proposed method for few representative frames.

From Fig. 1, it is observed that the proposed method recognizes different pair of activities accurately. These activities are walk-run, bend-jump and jack-skip. The recognition results of the proposed method for these activities are shown in Fig. 1(a–f). We have compared advantages and disadvantages of the proposed method with a selected number of other methods like moment [15], LBP [24], CSLBP [27] and HOG [20] based methods in terms of average accuracy, shown in Tables 1, 2, and 3.

On observing Tables 1, 2 and 3, it is clear that the average accuracy rate of the proposed method is the highest when compared to Moment [15], LBP [24], CSLBP [27] and HOG [20] based methods. Therefore, it is said that that the performance of the proposed method is better than other methods.

6 Conclusions

In this paper, we have discussed and demonstrated a classification based approach for human activity recognition. We have used novel integrated moment-local binary patterns feature descriptor and binary support vector machine (SVM) classifier. The experiments have been performed over Weizmann activity video dataset. Six activities have been considered for experimental evaluations. Among these six activities, pairs of two activities have been taken at a time for classification with binary SVM classifier. These are walk-run, bend-jump, and jack-skip pairs of activities. Performance of the proposed method has been demonstrated qualitatively and quantitatively. Quantitative results show that the proposed method outperforms other Moment [15], Local Binary Patterns [24], Center Symmetric Local Binary Patterns [27] and Histogram of Oriented Gradients [20] based methods.

Acknowledgment. This work was supported by the University Grants Commission, New Delhi, India under Grant No. F.No.36-246/2008(SR) and Council of Scientific and Industrial Research, Human Resource Development Group, India via grant no. 09/001/(0362)/2012/EMR-I.

References

1. Zhu, Y., Nayak, N.M., Roy-Chowdhury, A.K.: Context aware activity recognition and anomaly detection in video. IEEE J. Sel. Top. Signal Process. **7**(1), 91–101 (2013)
2. Chihani, B., Bertin, E., Jeanne, F., Crespi, N.: Context-aware systems: a case study. In: Cherifi, H., Zain, J.M., El-Qawasmeh, E. (eds.) DICTAP 2011, Part II. CCIS, vol. 167, pp. 718–732. Springer, Heidelberg (2011)
3. OPPORTUNITY Activity Challenge – http://www.opportunity-project.eu/node/58 (2013). Last accessed 27 July 2013
4. NOKIA Mobile Data Challenge – http://research.nokia.com/page/12000 (2013). Last accessed 27 July 2013
5. Khare, M., Srivastava, R.K., Khare, A.: Moving object segmentation in Daubechies complex wavelet domain. Accepted for publication in Signal Image and Video Processing. ISSN: 1863-1711. Springer (2013). doi:10.1007/s11760-013-0496-4
6. Nigam, S., Deb, K., Khare, A.: Moment invariants based object recognition for different pose and appearances in real scenes. In: 2nd IEEE International Conference on Informatics, Electronics & Vision (ICIEV*)*, pp. 1–5, Dhaka, Bangladesh, May 2013
7. Binh, N.T., Khare, A.: Object tracking of video sequences in curvelet domain. Int. J. Image Graph. **11**(1), 1–20 (2011)
8. Aggarwal, J.K., Ryoo, M.S.: Human activity analysis: a review. ACM Comput. Surv. **43**(3), article 16. http://cvrc.ece.utexas.edu/aggarwaljk/Publications/review_ryoo_hdr.pdf (2011)

9. Alemdar, H., van Kasteren, T.L.M., Ersoy, C.: Activity recognition with hidden Markov models using active learning. In: 19th IEEE Conference on Signal Processing and Communications Applications (SIU), pp. 1161–1164, Antalya (2011)
10. Cao, H., Nguyen, M.N., Phua, C., Krishnaswamy, S., Xiao, L.L.: An integrated framework for human activity classification. In: ACM Conference on Ubiquitous Computing (UbiComp'12), pp. 331–340, New York (2012)
11. Gorelick, L., Blank, M., Shechtman, E., Irani, M., Basri, R.: Actions as space time shapes. IEEE Trans. Pattern Anal. Mach. Intell. **29**(12), 2247–2253 (2007)
12. Liu, C., Yuen, J., Torralba, A.: SIFT flow: dense correspondence across scenes and its applications. IEEE Trans. Pattern Anal. Mach. Intell. **33**(5), 978–994 (2011)
13. Scovanner, P., Ali, S., Shah, M.: A 3-dimensional sift descriptor and its application to action recognition. In: ACM 15th International Conference on Multimedia, pp. 357–360 (2007)
14. Zhao, Y., Belkasim, S.: Multiresolution Fourier descriptors for multiresolution shape analysis. IEEE Signal Process. Lett. **19**(10), 692–695 (2012)
15. Flusser, J., Suk, T., Zitova, B.: Moments and Moment Invariants in Pattern Recognition. Wiley, New York (2009)
16. Hu, M.K.: Visual pattern recognition by moment invariants. IRE Trans. Inf. Theory **8**(2), 179–187 (1962)
17. Flusser, J., Suk, T.: Combined blur and affine moment invariants and their use in pattern recognition. Pattern Recogn. **36**(12), 2895–2907 (2003)
18. Chen, B.J., Shu, H.Z., Zhang, H., Chen, G., Toumoulin, C., Dillenseger, J.L., Luo, L.M.: Quaternion Zernike moments and their invariants for color image analysis and object recognition. Signal Process. **92**(2), 308–318 (2012)
19. Hosny, K.M.: Refined translation and scale Legendre moment invariants. Pattern Recogn. Lett. **31**(7), 533–538 (2010)
20. Dalal, N., Triggs, B.: Histogram of oriented gradients for human detection. In: IEEE International Conference on Computer Vision and Pattern Recognition, pp. 886–893, San Diego, CA, USA (2005)
21. Pang, Y., Yuan, Y., Li, X., Pan, J.: Efficient HOG human detection. Signal Process. **91**(4), 773–781 (2011)
22. Skibbe, H., Reisert, M., Schmidt, T., Brox, T.: Fast rotation invariant 3D feature computation utilizing efficient local neighborhood operators. IEEE Trans. Pattern Anal. Mach. Intell. **34**(8), 1563–1575 (2012)
23. Nigam, S., Khare, M., Srivastava, R.K., Khare, A.: An effective local feature descriptor for object detection in real scenes. In: IEEE Conference on Information and Communication Technologies (ICT 2013), pp. 244–248, Kanyakumari, India, April 2013
24. Pietikäinen, M., Hadid, A., Zhao, G., Ahonen, T.: Computer Vision using Local Binary Patterns. Computational Imaging and Vision, vol. 40. Springer, Heidelberg (2011)
25. Mu, Y., Yan, S., Liu, Y., Huang, T., Zhou, B.: Discriminative local binary patterns for human detection in personal album. In: IEEE International Conference on Computer Vision and Pattern Recognition, pp. 1–8, Anchorage, AK, (2008)
26. Nguyen, D.T., Ogunbona, P.O., Li, W.: A novel shape-based non-redundant local binary pattern descriptor for object detection. Pattern Recogn. **46**(5), 1485–1500 (2013)
27. Ko, B.C., Kim, D.Y., Jung, J.H., Nam, J.Y.: Three-level cascade of random forests for rapid human detection. Opt. Eng. **52**(2), 027204 (2013)
28. Shen, J., Yang, W., Sun, C.: Real-time human detection based on gentle MILBoost with variable granularity HOG-CSLBP. Neural Comput. Appl. (2012). doi:10.1007/s00521-012-1153-5