Recognition of Human Activities in Daubechies Complex Wavelet Domain

Manish Khare(✉)

Dhirubhai Ambani Institute of Information and Communication Technology
(DA-IICT), Gandhinagar, India
mkharejk@gmail.com

Abstract. Recognition of accurate human activities is a challenging research problem in video surveillance problem of computer vision research. The task of recognizing activities of human from video sequence exhibits more challenges because of real time processing of data. In this paper, we have proposed a method for recognition of human activities based on Daubechies complex wavelet transform (DCxWT). Better edge representation and approximate shift invariant properties of DCxWT over the other real valued wavelet transform motivates us to utilize properties of DCxWT in recognition of human activities. The multi-class SVM is used for classifying the recognized human activities. The proposed method is compared with other state-of-the-art method, on various standard publicly available dataset, in terms of different quantitative performance measures. We found that the proposed method has better recognition accuracy in comparison to other state-of-the-art methods.

Keywords: Human activity recognition · Daubechies complex wavelet transform · Multiclass support vector machine · Feature extraction

1 Introduction

Recognition of human activity is crucial and popular area in computer vision research, because it is the foundation of development of many applications [1, 2]. Human monitoring-based surveillance system has many industrial applications [3]. These types of system are also very useful in police investigation after a crime or any other illegal action has occurred. Hence, one can say that, human activity recognition-based system is very essential for effective video surveillance system. Clutter background, multi-view point, varying lighting condition, and other similar problems, make accurate and efficient human activity recognition a challenging task. The objective of human activity recognition is to automatically analyze ongoing activities from a video (sequence of frames), and classify an activity which is a member of a given set of abstract activities into one such abstract activity, for ex. Running, walking, jumping, etc.

Based on different detail studies [1, 2, 4, 5], human activity recognition approach can be classified in various categories. According to analysis done by Aggarwal and Ryoo [2], human activities can be conceptually categorized into four levels, depending on their complexities: gesture, actions, interaction, and group activities. Gesture are

© Springer Nature Switzerland AG 2019
E. Ricci et al. (Eds.): ICIAP 2019, LNCS 11752, pp. 357–366, 2019.
https://doi.org/10.1007/978-3-030-30645-8_33

elementary movements of body part of a person for ex. ‘raising a leg’ or ‘stretching an arm’. Actions are single person activity that are results of multiple gestures temporally such as ‘walking’, running’, ‘jogging’, etc. In Interaction type of human activity, two or more persons and/or objects are involved, for ex. ‘two-person fighting’ is an interaction between two humans. Group activities have more complexity is comparison to other three types. In group activities, activities are performed by a one group, which is composed of multiple persons and/or objects. For ex. ‘A group of persons marching’ is a typical example of group activities.

Now a day’s machine learning based approach for human activity recognition is widely used because the amount of unannotated training data is readily available for the unsupervised training of these systems. In this type of approach, after learning from a collection of data, algorithm try to answer questions related to that data. The training used in this type of approach is either pixel-based or feature-based [6]. Feature-based approach is better than the pixel-based approach in terms of execution time and speed. For any good computer vision application, an activity recognition algorithm should hold two properties - (i). activity recognition algorithm should perform under real time constraint, and (ii). activity recognition algorithm should be able to solve multi-view as well as multiclass problem.

A lot of works have been proposed to solve activity recognition problem in last few decades. A method for activity recognition using Principal Component Analysis, and Hidden Markov Model was proposed by Uddin et al. [7]. Method based on background subtraction with shape and motion information features for activity recognition method was proposed by Qian et al. [8] for a smart surveillance system. Bobick and Davis [9] proposed an activity recognition method based on motion templates. Human recognition method based on constrained Delaunay triangulation technique, which divides the posture into different triangular meshes, was proposed by Hsieh et al. [10]. Holte et al. [11] proposed a machine learning-based approach to detect motion of the actors by computing the optical flow in video data.

While most existing human action recognition methods adopted feature(s) that works for single resolution of images, an image can be with complex structures and consists of varying levels of details. To remedy this issue, multi-resolution analysis (MRA) can be adopted. In MRA, images can be analyzed at more than one resolution, so that the features that are left undetected at one level can get a chance to consider in another level. Wavelet transform is the most popular tool of MRA. Wavelet transform can be categorized into real valued and complex valued wavelet transforms. Real valued wavelet transform uses real-valued filters to get real valued coefficients while complex wavelet transform uses complex valued filters to get complex valued coefficients.

Method proposed by Khare et al. [12] uses real valued wavelet transform for activity recognition, but real valued wavelet transform is not suitable in various computer vision application. Use of complex wavelet transform can avoid different shortcomings of real valued wavelet transform. Khare et al. [13] proposed dual tree complex wavelet transform based approach for human action recognition. Dual tree complex wavelet transform is not a true sense complex wavelet transform and its implementation is based on real valued wavelet transform.

Motivated by the work of Khare et al. [12, 13], in this paper, we have proposed Daubechies complex wavelet transform (DCxWT) based approach for recognition of human activities. DCxWT is a true sense complex wavelet transform and having advantages of approximate shift-invariance and better edge representation as compared to real valued wavelet transform. We have used multiclass support vector machine as a classifier for classifying different human activities in video frames. We have experimented the proposed method at multiple levels of DCxWT and shown that performance of the proposed method is becoming better as we move toward higher levels. We have conducted the experiments on different standard action datasets for evaluation of the proposed method. The proposed method is compared with some other state-of-the-art methods proposed by Qian et al. [8], Holte et al. [11], and Khare et al. [12, 13], for showing effectiveness of DCxWT.

The rest of paper is organized as follows: Sect. 2 describes DCxWT used as a feature for recognition of human activities. Section 3 describes the proposed method in detail. Experimental results of the proposed method and other state-of-the-art methods are given in Sect. 4. Finally, conclusions of the work are given in Sect. 5.

2 Daubechies Complex Wavelet Transform

In any recognition algorithm, selection of appropriate feature is very important. If correct feature is selected for recognition then performance of classifier will improve. In our proposed work for recognition of human activities, we have used DCxWT coefficients as feature set. A brief description of Daubechies complex wavelet transform and why this is useful for recognition of human activities is given in below –

For activity recognition, we require a feature which remains invariant by shift, translation and rotation of object, because object may be present in translated and rotated form among different scenes. Due to its approximate shift-invariance and better edge representation property, we have used DCxWT as a feature for recognition of human activity.

Any function $f(t)$ can be decomposed into complex scaling function and mother wavelet as:

$$f(t) = \sum_{k} c_k^{j_0} \phi_{j_0,k}(t) + \sum_{j=j_0}^{j_{\max}-1} d_k^j \psi_{j,k}(t) \tag{1}$$

where j_0 is given low resolution level, $\left\{c_k^{j_0}\right\}$ and $\left\{d_k^j\right\}$ are approximation coefficients $\left[\phi(u) = 2\sum_i a_i \phi(2u - i)\right]$ and detail coefficients $\left[\psi(t) = 2\sum_n (-1)^n \overline{a_{1-n}} \phi(2t - n)\right]$. where $\phi(t)$ and $\psi(t)$ share same compact support [−L, L + 1] and a_i's are coefficients. The a_i's can be real as well as complex valued and $\sum a_i = 1$.

Daubechies's wavelet bases $\{\psi_{j,k}(t)\}$ in one-dimension is defined through above scaling function $\phi(u)$ and multiresolution analysis of $L_2(\Re)$[14]. During the formulation of general solution if we relax the condition for a_i to be real [14], it leads to complex valued scaling function.

DCxWT holds various properties [14], in which reduced shift sensitivity and better edge representation properties of DCxWT are important one for classification

3 The Proposed Method

This section describes, new method for human activity recognition, in which we used Daubechies complex wavelet transform coefficients as a feature of objects. Block diagram of the proposed method is given in following Fig. 1.

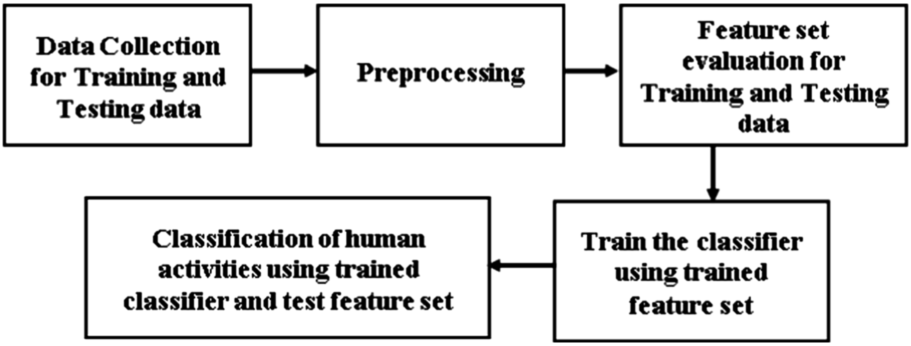
Data Collection for Training and Testing data
Preprocessing
Feature set evaluation for Training and Testing data
Train the classifier using trained feature set
Classification of human activities using trained classifier and test feature set

Fig. 1. Block diagram of the proposed method

Details of the proposed method are described in following steps –

Step 1: first step of the proposed method is to collect and represent data for training and testing. In our method, we have tested the proposed method with various publicly available standard datasets for ex. CASIA dataset [15] and KTH dataset [16]. Human activities are presented in form of video which is a sequence of frames. These videos can be used either for training or testing, depends on purpose. Each frame of video can be considered as an image.

Step 2: Videos in dataset, are may be of different size or having different color format. Therefore, normalization of these videos is required to reduce complexity and maintain uniformity between all dataset with respect to algorithm. In the present work, the collected data are scale normalized to 256 × 256-pixel dimension. After scale normalization, color format of normalized collected data was converted into gray-level format.

Step 3: Third step of the proposed method is feature set computation. For feature set computation in the proposed method, frames of video dataset are decomposed into complex wavelet coefficients using DCxWT. After applying DCxWT, coefficients are in form of four sub-bands namely – LL, HL, LH, HH, in which LL is approximately coefficients and rest other HL, LH, HH gives details coefficients. Each detail coefficient matrix (HL, LH, and HH) is used separately to construct feature vector. Values of LL sub-bands are dropped for construction of feature and this is used for further higher-level decomposition due to multi-resolution property of DCxWT. When we decompose wavelet transform for level 2, then in this level, we further decompose approximation coefficients (LL sub-band) of level 1. This again produces one approximation coefficients and three detail coefficients matrix. In case of level 2, the feature of level 1 are combined with the present level 2, which constitute level 2. Similar process is repeated for successive level of decomposition.

Step 4: Forth and last step of the proposed method is recognition of human activities. For this process, we have performed classification-based recognition of human activity. For this process, we compute feature set of test and trained video dataset using step 1–3 of the proposed method, then both trained and test feature sets are supplied into multi-class SVM classifier. Multi-class SVM classifier analyzes test feature set with trained feature set and gives result in form of recognized human activities. Same process will repeat for all other video datasets. In multi-class SVM classifier we have used radial basis function (RBF) as a kernel function.

4 Experimental Results

In this section, we demonstrate the experimental results of the proposed method, with those of the method proposed by Qian et al. [8], Holte et al. [11], and Khare et al. [12, 13], in terms of Precision, Recall, F-Score, and Recognition Accuracy. The proposed method and other state-of-the-art methods, mentioned above for human activity recognition have been presented here for CASIA action datasets [15] and KTH dataset [16].

CASIA dataset is a collection of sequences of human activities captured by video cameras in outdoor environment from different angle of view. In CASIA dataset, video sequences have non-uniform background. In this dataset, a total of 1446 video sequences, containing fifteen types of different actions. All video sequences were taken simultaneously with three non-calibrated cameras from different view angles (horizontal view, angle view, and top-down view). Resolution of videos in dataset is 320×240 pixels with frame rate 25 frames per second. Figure 2, shows the different activities with different viewpoints of CASIA dataset, in which we have performed experiments. For the experiments, we have taken five different activities out of 15 activities, namely – walk, run, bend, fight, and rob. KTH dataset [16] is one of the largest dataset with 192 videos for training, 192 videos for validation, and 216 videos for testing, in which six types of human actions (Walking, Jogging, Running, Boxing, Hand waving, and Hand clapping), performed by 25 persons. Figure 3 shows the different activities of KTH dataset [16].

From the Figs. 2 and 3, one can see that, human activities are differently illuminated and objects are scaled in different camera positions as well as one can observe that both frontal as well as side view of objects are taken for experimentation of human action recognition.

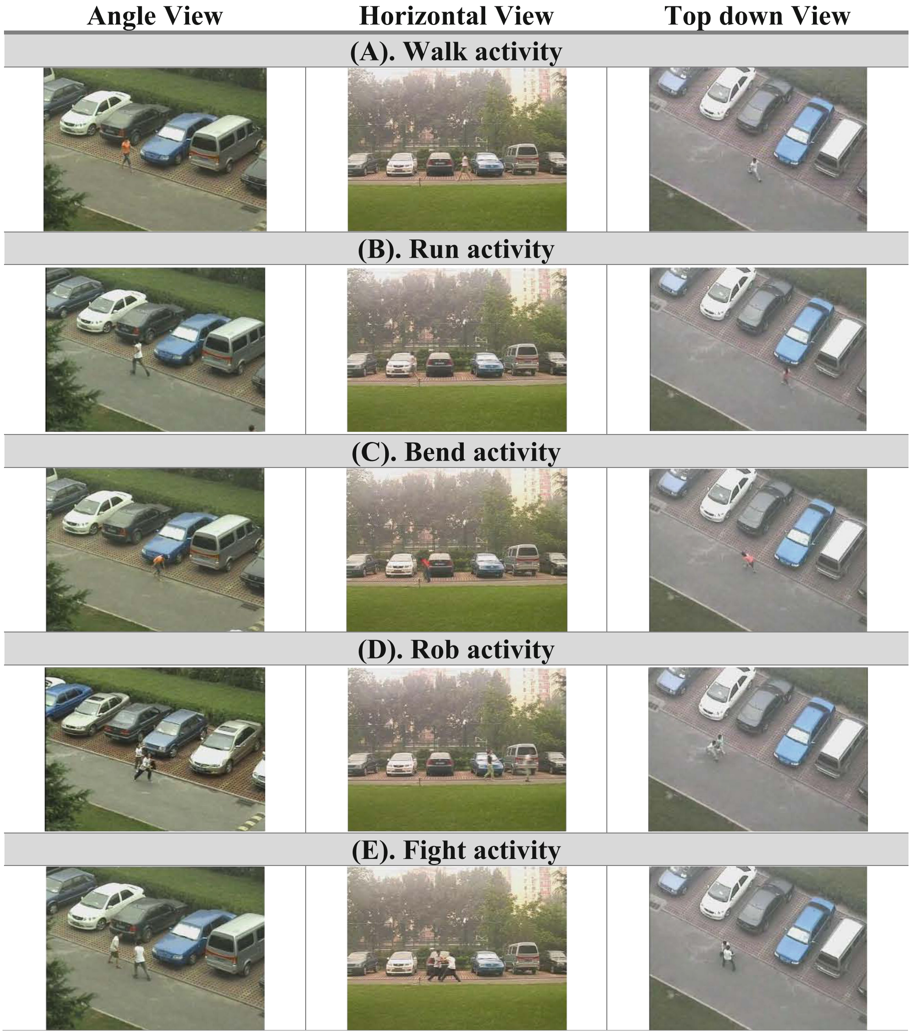

Fig. 2. Examples of different activities in the CASIA Dataset [15]

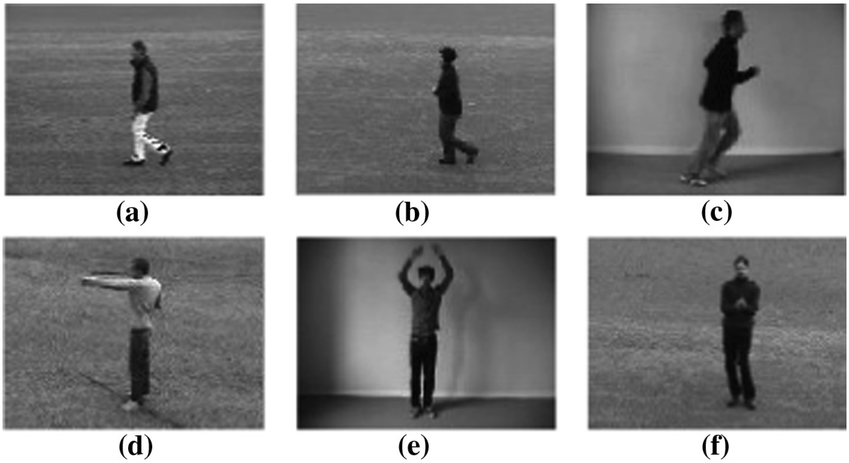

Fig. 3. Examples of KTH Dataset frames for different activities [16] [(a) Walking, (b) Jogging, (c) Running, (d) Boxing, (e) Hand waving, (f) Hand clapping]

Recognition results for CASIA dataset [15] and KTH dataset respectively are shown in Tables 1 and 2 for the proposed method and other state-of-the-art methods [8, 11–13]. For comparison we have used four different performance metrices [17] including Precision, Recall, F-score and Recognition accuracy.

Precision is an average per-class agreement of the data class labels with those of a classifier. Recall is an average per-class effectiveness of a classifier to identify class labels. F-Score is a relation between data positive labels and those given by a classifier based on a per-class average; it conveys the balance between the precision and recall. Recognition accuracy is an average per-class effectiveness of a classifier.

$$Precision = \frac{\sum_{i=1}^{L}\left(\frac{TP_i}{TP_i + FP_i}\right)}{L} \tag{2}$$

$$Recall = \frac{\sum_{i=1}^{L}\left(\frac{TP_i}{TP_i + FN_i}\right)}{L} \tag{3}$$

$$F{-}Score = 2.\left(\frac{Precision * Recall}{Precision + Recall}\right) \tag{4}$$

$$Accuracy = \frac{\sum_{i=1}^{L}\left(\frac{TP_i + TN_i}{TP_i + FN_i + FP_i + TN_i}\right)}{L} \tag{5}$$

where L is the number of classes, TP is true positive, FP is false positive, FN is false negative, and TN is true negative.

Tables 1 and 2, shows the recognition results in form of four different performance evaluation matrices for the proposed method and other state-of-the art methods [8, 11–13] for CASIA dataset [15] and KTH dataset [16] respectively. From Tables 1 and 2, we can see that, the proposed method gives better recognition accuracy in comparison to other state-of-art methods [8, 11–13]. Methods proposed by Holte et al. [11] works for multi-view human activity recognition, but from table we can see that the proposed method gives better recognition results in comparison to these methods, so that we can see that the proposed method works good for multi-view poses as well.

Table 1. Performance measures values for CASIA Dataset [15]

Method		Precision	Recall	F-score	Recognition accuracy
The proposed method with	DCxWT (Level - 1) as a feature	0.8339	0.8240	0.8257	0.9295
	DCxWT (Level - 3) as a feature	0.8606	0.8520	0.8534	0.9408
	DCxWT (Level - 5) as a feature	0.8852	0.8780	0.8789	0.9512
	DCxWT (Level - 7) as a feature	0.8981	0.8940	0.8938	0.9572
Method with DTCWT as feature (Khare et al. [13])	DTCWT (Level - 1) as a feature	0.8008	0.7980	0.7985	0.9192
	DTCWT (Level - 3) as a feature	0.8221	0.8200	0.8203	0.9280
	DTCWT (Level - 5) as a feature	0.8408	0.8380	0.8385	0.9352
	DTCWT (Level - 7) as a feature	0.8610	0.8580	0.8584	0.9430
Method with DWT as a feature (Khare et al. [12])	DWT (Level - 1) as a feature	0.7544	0.7440	0.7456	0.8976
	DWT (Level - 3) as a feature	0.7761	0.7660	0.7682	0.9068
	DWT (Level - 5) as a feature	0.7899	0.7820	0.7835	0.9128
	DWT (Level - 7) as a feature	0.8164	0.7791	0.7964	0.9181
Qian et al. [8]		0.7719	0.7640	0.7658	0.9056
Holte et al. [11]		0.8189	0.8180	0.8182	0.9272

Table 2. Performance measures values for KTH Dataset [16]

Method		Precision	Recall	F-score	Recognition accuracy
The proposed method with	DCxWT (Level - 1) as a feature	0.8210	0.8117	0.8148	0.9378
	DCxWT (Level - 3) as a feature	0.8277	0.8317	0.8333	0.9439
	DCxWT (Level - 5) as a feature	0.8632	0.8583	0.8598	0.9528
	DCxWT (Level - 7) as a feature	0.9013	0.8983	0.8993	0.9661
Method with DTCWT as feature (Khare et al. [13])	DTCWT (Level - 1) as a feature	0.8008	0.7933	0.7946	0.9311
	DTCWT (Level - 3) as a feature	0.8216	0.8283	0.8227	0.9400
	DTCWT (Level - 5) as a feature	0.8573	0.8517	0.8525	0.9506
	DTCWT (Level - 7) as a feature	0.8755	0.8716	0.8724	0.9572
Method with DWT as a feature (Khare et al. [12])	DWT (Level - 1) as a feature	0.7320	0.7284	0.7292	0.9094
	DWT (Level - 3) as a feature	0.7733	0.7567	0.7599	0.9189
	DWT (Level - 5) as a feature	0.7838	0.7700	0.7728	0.9233
	DWT (Level - 7) as a feature	0.7854	0.7717	0.7745	0.9239
Qian et al. [8]		0.8252	0.8183	0.8200	0.8394
Holte et al. [11]		0.9445	0.9350	0.9391	0.9297

5 Conclusion

In this paper, we demonstrated a new method for recognition of human activities. The proposed approach used DCxWT coefficients as a feature of activities of human objects. We evaluated results for multiple levels of DCxWT. we compared results of the proposed method with other state-of-the-art methods [8, 11–13] on CASIA dataset [15] and KTH dataset [16], in terms of four different performance measures: Precision, Recall, F-Score, and Recognition accuracy. From the results, we could conclude that the proposed method for human action recognition has better recognition accuracy at higher levels of DCxWT in comparison to DWT or DTCWT, and the proposed method outperforms the other methods. From the results, we can conclude that the complex wavelet transform is better than real wavelet transforms for action recognition in terms of the different performance measures.

References

1. Vrigkas, M., Nikou, C., Kakadiaris, I.A.: A review of human activity recognition methods. Frontiers Robot. AI **2**, 28 (2015)
2. Aggarwal, J.K., Ryoo, M.S.: Human activity analysis: a review. ACM Comput. Surv. **43**(3), 15 (2011)
3. Collins, R.T., Lipton, A.J., Kanade, T.: Introduction to the special section on video surveillance. IEEE Trans. Pattern Anal. Mach. Intell. **22**(8), 745–746 (2000)
4. Ziaeefard, M., Bergevin, R.: Semantic human activity recognition: a literature review. Pattern Recogn. **48**(8), 2329–2345 (2015)
5. Borges, P.V.K., Conci, N., Cavallaro, A.: Video-based human behavior understanding: a survey. IEEE Trans. Circuits Syst. Video Technol. **23**(11), 1993–2008 (2013)
6. Moeslund, T.B., Hilton, A., Kruger, V.: A survey of advances in vision-based human motion capture and analysis. Comput. Vis. Image Underst. **104**(2–3), 90–126 (2006)
7. Uddin, M.Z., Lee, J.J., Kim, T.S.: Independent shape component-based human activity recognition via hidden markov model. Appl. Intell. **33**(2), 193–206 (2010)
8. Qian, H., Mao, Y., Xiang, W., Wang, Z.: Recognition of human activities using SVM multi-class classifier. Pattern Recogn. Lett. **31**(2), 100–111 (2010)
9. Bobick, A.F., Davis, J.W.: The recognition of human movement using temporal templates. IEEE Trans. Pattern Anal. Mach. Intell. **23**(3), 257–267 (2001)
10. Hsieh, J.W., Hsu, Y.T., Liao, H.Y.M., Chen, C.C.: Video-based human movement analysis and its application to surveillance systems. IEEE Trans. Multimed. **10**(3), 372–384 (2008)
11. Holte, M.B., Moeslund, T.B., Nikolaidis, N., Pitas, I.: 3D human action recognition for multi-view camera systems. In: Proceeding of International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission, pp. 342–349 (2011)
12. Khare, M., Jeon, M.: Towards discrete wavelet transform based human activity recognition. In: Proceeding of 2nd International Workshop on Pattern Recognition (IWPR 2017), p. 1044308 (1-5), Singapore (2017)
13. Khare, M., Gwak, J., Jeon, M.: Complex wavelet transform-based approach for human action recognition in video. In: Proceeding of International Conference on Control, Automation and Information Sciences (ICCAIS 2017), pp. 157–162, Thailand (2017)
14. Clonda, D., Lina, J.M., Goulard, B.: Complex daubechies wavelets: properties and statistical image modeling. Sig. Process. **84**(1), 1–23 (2004)
15. Wang, Y., Huang, K., Tan, T.: Human activity recognition based on R transform. In: Proceedings of International Conference Computer Vision and Pattern Recognition (CVPR 2007), pp. 1–7 (2007). http://www.cbsr.ia.ac.cn/english/Action%20Databases%20EN.asp
16. Schuldt, C., Laptev, I., Caputo, B.: Recognizing human actions: a local SVM approach. In: Proceedings of 17th International Conference on Pattern Recognition, vol. 3, pp. 32–36 (2004). http://www.nada.kth.se/cvap/actions/
17. Sokolove, M., Lapalme, G.: A systematic analysis of performance measures for classification tasks. Inf. Process. Manage. **45**, 427–437 (2009)