Scale Invariant Action Recognition Using Compound Features Mined from Dense Spatio-temporal Corners

Andrew Gilbert, John Illingworth, and Richard Bowden

CVSSP, University of Surrey, Guildford,
GU2 7XH, England

Abstract. The use of sparse invariant features to recognise classes of actions or objects has become common in the literature. However, features are often "engineered" to be both sparse and invariant to transformation and it is assumed that they provide the greatest discriminative information. To tackle activity recognition, we propose learning compound features that are assembled from simple 2D corners in both space and time. Each corner is encoded in relation to its neighbours and from an over complete set (in excess of 1 million possible features), compound features are extracted using data mining. The final classifier, consisting of sets of compound features, can then be applied to recognise and localise an activity in real-time while providing superior performance to other state-of-the-art approaches (including those based upon sparse feature detectors). Furthermore, the approach requires only weak supervision in the form of class labels for each training sequence. No ground truth position or temporal alignment is required during training.

1 Introduction

The recognition of human activity within a video sequence is a popular problem. It is a difficult as subjects can vary in size, appearance and pose. Furthermore, cluttered backgrounds and occlusion can also cause methods to fail. Varying illumination and incorrect temporal alignment of actions can cause large within (intra) class variation. While inter-class variation can be low due to similarity in motion and appearance. To illustrate, Figure 3(d), (e) & (f) show example frames from the KTH [1] dataset for the categories 'jogging', 'running' and 'walking' respectively. Scaling issues aside, the similarity of these static frames illustrates the need to use temporal information when identifying actions.

Within the object recognition community, learning strategies for feature selection have proven themselves successful at building classifiers from large sets of possible features e.g. Boosting [2]. Although similar approaches have been applied to the spatio-temporal activity domain [3] [4], such approaches do not scale well due to the number of features and also issues with time alignment/scaling. Therefore sparse, but more complex, feature descriptors have been proposed [5] [1] [6]. The sparsity of such features makes the problem of recognition tractable but such sparsity also means potential information is lost to the recognition architecture.

D. Forsyth, P. Torr, and A. Zisserman (Eds.): ECCV 2008, Part I, LNCS 5302, pp. 222–233, 2008.

Our approach is based upon extracting very low-level features (corners in xy, xt and yt) from videos and combining them locally to form high-level, compound, spatio-temporal features. The method outlined in this paper takes advantage of data mining to assemble the compound features using an *association rule* data mining technique [7] which efficiently discovers frequently reoccurring combinations/rules. The resulting rules are then used to form a classifier which provides a likelihood of the occurrence and position of an action in a sequence.

Association rule data mining was recently employed by Quack *et al.* [8] to group SIFT descriptors for object recognition. We use the algorithm in a similar fashion but instead of using it to group high-level features, we use it to build high-level compound features from a noisy and over-complete set of low-level spatio and spatio-temporal features (corners). This is then applied to the task of activity recognition. We compare encoding only relative spatial offsets, which provides scale invariance, to the spatial grid proposed by Quack *et al.* and demonstrate that, due to increased scale invariance, higher performance is achieved. Learning is performed with only sequence class labels rather than full spatio-temporal segmentation. The resulting classifier is capable of both recognising and localising activities in video. Furthermore, we demonstrate that efficient matching can be used to obtain real-time action recognition on video sequences.

2 Related Work

Within object recognition, the use of spatial information of local features has shown considerable success [8] [9] [10]. Many action recognition methods also use a sparse selection of local interest points. Schüldt *et al.* [1] and Dollar *et al.* [5] employ sparse spatio-temporal features for the recognition of human (and mice) actions. Schüldt takes a codebook and bag-of-words approach applied to single images to produce a histogram of informative words or features for each action. Niebles and Fei-Fei [11] use a hierarchical model that can be characterized as a constellation of bags-of-words. Similarly Dollar take the bag-of-words approach but argue for an even sparser sampling of the interest points. This improves the performance on the same video sets. However, with such a sparse set of points, the choice of feature used is important. Scovanner *et al.* [12] extended the 2D SIFT descriptor [13] into three dimensions, by adding a further dimension to the orientation histogram. This encodes temporal information and dramatically outperforms the 2D version. To model motion between frames, optical flow [14] [15] can be applied as was used by Laptev [6] in addition to a shape model to detect drinking and smoking actions. Yang Song *et al.* [16] use a triangular lattice of grouped point features to encode layout.

There are relatively few examples of mining applied to the imaging domain. Tesic *et al.* [17] use a Data mining approach to find the spatial associations between classes of texture from aerial photos. Similarly Ding *et al.* [18] derive association rules on Remote Sensed Imagery data using a Peano Count Tree (P-tree) structure with an extension of the more common *APriori* [7] algorithm.

Chum *et al.* [19] used data mining to find near duplicate images within a database of photographs. Our approach uses data mining as a feature selection process for activity recognition.

3 Building Compound Features

3.1 Extracting Temporal Harris Interest Points

In contrast to very sparse feature detectors, we build our detection system upon corner features. The rationale for using corners are they are simple to compute, largely invariant to both lighting and geometric transformation, and provide an over-complete feature set from which to build more complex compound features. To identify and locate the interest points in images, the well known Harris corner detector [20] is applied in (x, y), (x, t) and (y, t) as a 3x3 patch. Unlike the 3D corners of [6], which are sparse, detecting 2D corners in 3 planes produces a relatively large and over complete set of features, with typically 400 corners detected per frame on the KTH data. Each corner feature has a dominant gradient orientation, this orientation can be used to encode the feature type into one of a set of discrete corner orientations. Figure 1 shows the example corner detections on two frames. It shows that in 1(a), most features occur around the hands espe-

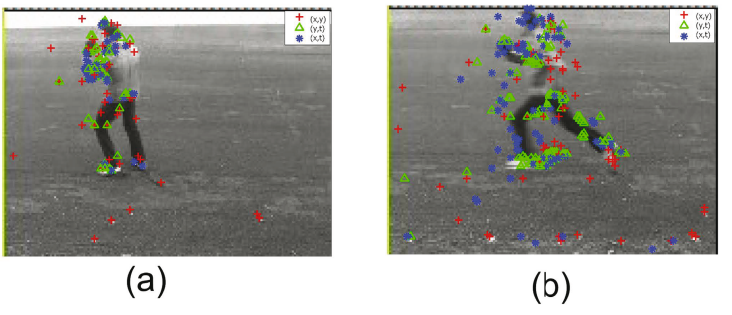

Fig. 1. Corner Detection on two Frames, (a) A Boxing Sequence, (b) A Running Sequence

cially in the (x,t) and (y,t) dimensions. A similar pattern occurs in 1(b) with a large amount of features around the feet, hands and head. The large number of features detected make clustering methods for code book construction unsuitable but the simplicity of the features also makes such an approach redundant.

In order to overcome the effects of scale, the interest point detector was applied to the video sequences across scale space to detect corners at different scales [21]. This was achieved by successively 2x2 block averaging the image frames. Table 1 shows the scale, image size and effective interest point patch sizes. Each feature is now encoded by a 3 digit vector (s, c, o). The encoding includes the scale $s \in \{1, ..., 4\}$ corresponding to the interest point size $\{3x3, ..., 48x48\}$, $c \in \{1, ..., 3\}$ the channel the interest point was detected in $\{xy, xt, yt\}$ and the

Table 1. Table showing the image and relative interest point patch sizes

Scale	1	2	3	4
Image Size	160x120	80x60	40x30	20x15
Interest Point Size	3x3	6x6	24x24	48x48

gradient orientation of the interest point $o \in \{1, ..., n\}$. Orientation is quantised into n discrete orientations. In our experiments $n = 8$ so orientation is quantised into 45^o bins aligned with a points of a compass. Figure 2(a) shows an example of the vector encoding.

3.2 Spatial Grouping

The spatial configuration of features is key to object recognition and has been demonstrated to significantly enhance action recognition when modelled independently from temporal information [6]. Quack *et al.* [8] encoded the spatial layout of features by quantising the space around a feature into a grid and assigning features to one of those locations. Where, the size of the grid is dependant on the scale of the detected SIFT feature to provide robustness to scale. This approach is difficult to achieve for less descriptive interest points such as corners, so our approach is to define neighbourhoods centred upon the feature that encode the relative displacement in terms of angle rather than distance hence achieving scale invariance. To do this, each detected interest point forms the centre of a neighbourhood. The neighbourhood is divided into 8 quadrants in the x, y, t domain which radiate from the centre of the neighbourhood out to the borders of the image in x, y and one frame either side either $t - 1$ or $t, t+1$ (see Figure2(b-c)). Each quadrant is given a label, all feature codes found

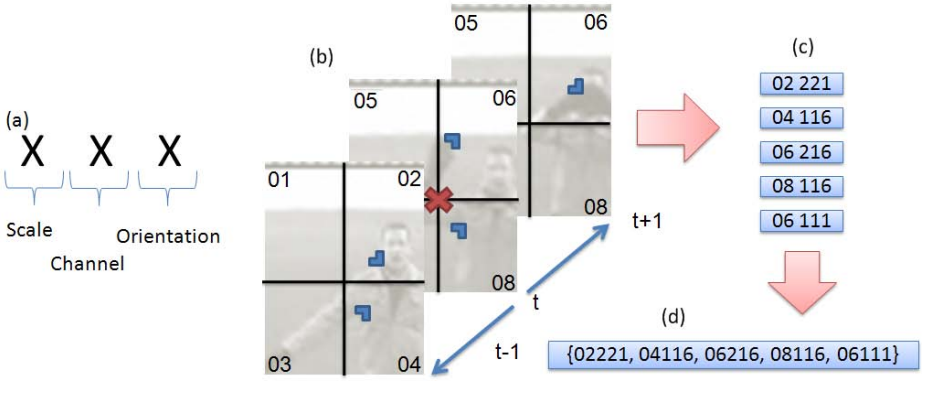

Fig. 2. (a) The three parts that make up a local feature descriptor. (b) shows a close-up example of a 2x2x2 neighbourhood of an interest point, with five local features shown as corners. (c) shows the spatial and temporal encoding applied to each local feature. (d) Concatenating the local features into a transaction vector for this interest point.

within a unique quadrant are appended with the quadrant label. A vector of these elements is formed for every interest point found in the video sequence and contains the relative spatial encoding to all other features on the frame. For efficiency this is done by using a look-up to an integral histogram of the 3 digit feature codes. This newly formed set is called a transaction set, T, where the spatially encoded features contained within it are items. To summarise, Figure 2 shows the formation of a single transaction set, from five individual local features.

For each interest point a transaction set is formed. These are collected together to compute a transaction database for each action. For a typical example video from the KTH dataset, this database contains around 500,000 transactions for each action, where a single transaction contains around 400 items. To condense or summarise this vast amount of information, data mining is employed.

4 Data Mining

Association rule [22] mining was originally developed for the analysis of customers supermarket baskets. Its purpose, to find regularity in the shopping behaviour of customers, by finding association rules within millions of transactions. An association rule is a relationship of the form $\mathbf{A} \Rightarrow \mathbf{C}$, where $\mathbf{A}$ and $\mathbf{C}$ are itemsets. $\mathbf{A}$ is called the antecedent and $\mathbf{C}$ the consequence. An example of the rule can be, customers who purchase an item in $\mathbf{A}$ are very likely to purchase another item in $\mathbf{C}$ at the same time. As there will be billions of transactions and therefore millions of possible association rules, efficient algorithms have been developed to quickly formulate the rules. One such algorithm is the popular *APriori* algorithm developed by Agrawal [7].

4.1 Frequent Itemsets

It can be said that transaction T *supports* an itemset $\mathbf{A}$ if $\mathbf{A} \subseteq T$. The algorithm attempts to find subsets which are frequent to at least a minimum number T_{Conf} (confidence threshold) of the items. If $\{\mathbf{A}, \mathbf{B}\}$ is a frequent itemset, both subsets $\mathbf{A}$ and $\mathbf{B}$ must be frequent itemsets as well. This fact is exploited by the algorithm to increase efficiency. *APriori* uses a "bottom up" approach, where frequent subsets are extended one item at a time, and groups of candidates are tested against the confidence threshold.

4.2 Association Rules

The association rule is the expression $\{\mathbf{A}, \mathbf{B}\} \Rightarrow \mathbf{C}$ where given itemsets $\mathbf{A}$ and $\mathbf{B}$, the itemset $\mathbf{C}$ will frequently occur. The belief of each rule is measured by a support and confidence value.

Support Rule. The support, $sup(\{\mathbf{A}, \mathbf{B}\} \Rightarrow \mathbf{C})$ of a rule, measures the statistical significance of a rule, the probability that a transaction contains itemsets $\mathbf{A}$ and $\mathbf{B}$.

Confidence Rule. The confidence rule is used to evaluate an association rule. The confidence of a rule $Conf(\{\mathbf{A}, \mathbf{B}\} \Rightarrow \mathbf{C})$ is the support of the set of all items that appear in the rule, divided by the support of the antecedent of the rule. This means the confidence of a rule is the number of times in which the rule is correct relative to the number of cases in which it is applicable. This measure is used to select association rules, if it's confidence exceeds a threshold T_{Conf}.

4.3 Mining for Frequent and Distinctive Itemsets

Once the local feature neighbourhoods are formed into transactions, the frequent and distinctive itemsets that make up the transactions must be found. This is achieved by running the APriori [7] algorithm on the transaction database, to find the frequently occurring itemset configurations. It is important that the resulting frequent itemsets are distinctive inter class. Therefore positive examples of an action transaction were appended with a 1. While an equal sub set of all other actions are appended with a 0 to provide the negative examples for training. This is used as it is important the resulting mined itemset configurations are only frequent in assigning a feature to an action. Given an association rule **AS**, its confidence is used to look for rules that have a high probability of being correct. Meaning that a chosen frequent itemset must imply the specific action, as shown in Equation 1.

$$Conf(\mathbf{AS} \Rightarrow action) > T_{Conf} \tag{1}$$

The mining algorithm allows for the efficient computation of frequent itemset configurations. In our experiments, a transaction file consists of over 1 million possible transactions with each individual transaction containing around 400 items. This size would prohibit many semi-unsupervised learning methods. However the efficient approach of the APriori algorithm, allows for the frequent itemsets to be found within 1 hour, on standard desktop PC. Once completed, each association rule, **AS**, which satisfies equation 1 is added to a Frequent Mined Configuration vector $\mathbf{M}$. Where $\mathbf{M} = \{\mathbf{AS}_1, ..., \mathbf{AS}_N\}$ for the N association rules.

5 Classifying Actions

The Frequent Mined Configurations $\mathbf{M}$ for a specific action represents the frequent and distinctive itemsets of the training action sequences. Given a new query action sequence, the same feature extraction and spatial grouping of section 3 is applied to the query video. This forms a new query set of transactions $\mathbf{D}_{query} = \{T_1, ..., T_n\}$. To classify the action, a global classifier is used. However, in practice the extraction process is not required as the transaction rules can be applied as a lookup to the integral histogram.

5.1 Global Classifier

As shown in equation 2, the global classifier exhaustively compares a specific action (α) itemset $\mathbf{M}_\alpha$ and the image feature combinations in the transaction set

$\mathbf{D}_{query}$ for a triplet of frames $\mathbf{F} = \{f_{t-1}, f_t, f_{t+1}\}$ within a test sequence. It works as a voting scheme by accumulating the occurrences of the mined compound features.

$$Conf_\alpha(\mathbf{F}) = \frac{1}{N_\alpha * n} \sum_{\forall \mathbf{D}_{query}} m(T_i, \mathbf{M}_\alpha) \tag{2}$$

where N_α is the number of transaction sets mined from the training data, and n is the number of transactions or neighbourhoods in the current time step. $m(T_i, \mathbf{M}_\alpha)$ describes if a transaction is present in the mined configuration.

$$m(T_i, \mathbf{M}_\alpha) = \begin{cases} Conf(T_i \Rightarrow \alpha) & T_i \in \mathbf{M}_\alpha \\ 0 & otherwise \end{cases} \tag{3}$$

This is repeated over the complete test sequence of an action with all the mined action configurations to find the likelihood of the sequence. A correct match will occur often in equation 3 as the mining will select frequently reoccurring items that are distinct to other actions. The use of a codebook allows the classifier to run at approximately 12fps on unoptimised C++ code on a standard pc. Each video sequence is then classified as the action, α, for which the votes are maximised.

5.2 Action Localisation

As each transaction encodes the relative location of features into one of eight quadrants. Each transaction found can vote for which of the eight quadrants other features should be located in. A comparison is made between the features in the transaction set $\mathbf{D}_{query}$, with the Frequent Mined Configuration vector features $\mathbf{M}$. If a match is found, all pixels within a quadrant are incremented by 1 on a likelihood image. This is repeated for all matched features, eventually causing the likelihood image to produce a peak around the centre of the action. An example of this is shown in Figure 6(f), where Figure 6(e) shows the thresholded centre of the action.

6 Experiments

To evaluate the approach, two sets of videos were used. The KTH human action dataset of Schüldt *et al.* [1] is a popular dataset for action recognition, containing 6 different actions; boxing, hand-waving, hand-clapping, jogging, running and walking. There are a total of 25 people performing each action 4 times, giving 599 videos, (1 is missing) totalling 2396 unique actions. The portion of data for training and testing was identical to that proposed by Schüldt [1] to allow direct comparison of results. In order to demonstrate localisation in the presence of multiple subjects, a sequence consisting of a two people walking through the scene, with one person stopping to perform a single hand wave was recorded. Examples of the two sequences are shown in Figure 3. The sequences have different scales, and temporal speeds of actions, and some of the action classes

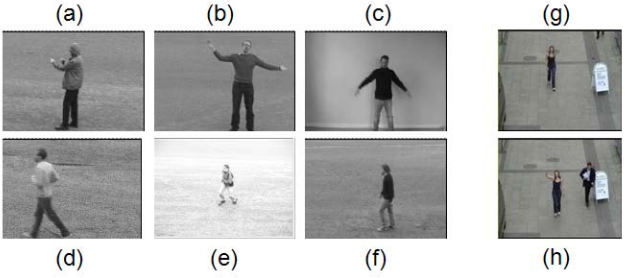

Fig. 3. Example frames from the two datasets, (a-f) KTH, (g,h) multi-person dataset: (a) boxing, (b) hand-clapping, (c) hand-waving, (d) jogging, (e) running, (f) walking, (g) one person walking, (h) one person walking, one person hand-waving

have very similar appearances. The training sequences, were used to produce a Frequent Mined Configuration vector M for each of the six actions containing up to 10 compound features in length. These were then used to classify each of the test sequences. Figure 4(a) shows the classification confusion matrix using the scale invariant grid approach proposed within this paper, where good class separability is exhibited. The results show relatively little confusion compared to other approaches with minor confusion between boxing and clapping. Jogging and running also causes some confusion but, this is consistent with previous approaches. This confusion is due to the inherent similarity of the motion. In Figure 4(b) the experiments were repeated using a *fixed size* 4x4 grid similar to [8]. To investigate the importance of the spatial and temporal compounding of individual features, Figure 5 shows the effect on overall accuracy (left axis) as the minimum item size in the transaction sets is increased. It can be seen

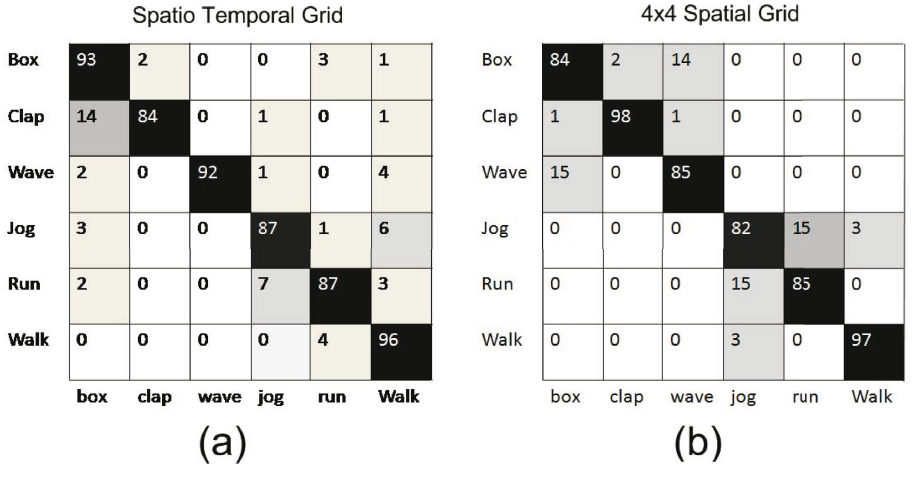

Fig. 4. (a)The confusion matrix of the Data Mined corner descriptor on the KTH dataset with **Scale Invariance**. (b) The confusion matrix of the Data Mined corner descriptor on the KTH dataset with a fixed non scale invariant spatial grouping.

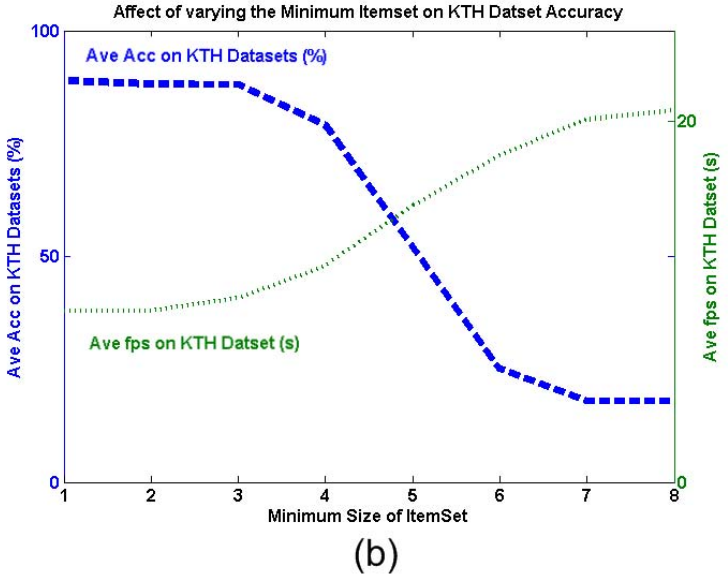

Fig. 5. The classification accuracy as the itemset size is increased

that no drop in performance is found in discarding itemsets under fours features in size. This confirms the importance of the grouping of the single features together. Disregarding these features gives an increase in frame rate from 9.5fps to 12fps, due to the reduced feature complexity. Therefore the small feature groups can be discarded with no loss in accuracy to further increase speed.

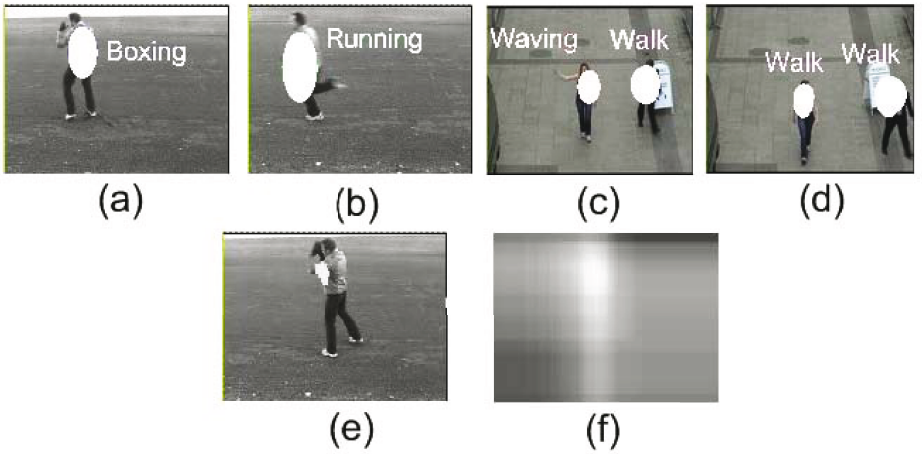

Fig. 6. (a)Localised boxing action (b) Localised running action Likelihood image, (c) Multiple localised waving and walking actions, (d) Multiple localised walking actions (e) Thresholded localised action (f) Localisation likelihood image for image (e).

The classification can also be used to localise and identify multiple actions in frames. Figure 6 shows the localisation of four frames actions. Two are from the KTH sequences (a) and (b), while (c) and (d) are from the multi-person outdoor sequence, it contains two people, walking where one stops and waves. In addition the wave action is much less exaggerated than the KTH version and only single handed. Despite these constraints, as shown in Figure 6(c) and (d), the actions are correctly localised and identified.

Table 2 shows results by a number of previous published works on the same dataset, including **Spat-Temp Dollar:** The very sparse spatio-temporal descriptor by Dollar [5] and **Subseq Boost Nowozin:** The boosted SVM classifier by Nowozin [23]. As Table 6 shows, our proposed technique, **Scale Invariant**

Table 2. Comparison of Average precision compared to other techniques on KTH action recognition Dataset

Method	Average Precision
Nowozin *et al.* [23] Subseq Boost SVM	87.04%
Wong and Cipolla [24] Subspace SVM	86.60%
Niebles *et al.* [25] pLSA model	81.50%
Dollar *et al.* [5] Spat-Temp	81.20%
Schüldt *et al.* [1] SVM Split	71.71%
Ke *et al.* [3] Vol Boost	62.97%
Fixed Grid Mined Dense Corners	88.50%
Scale Invariant Mined Dense Corners	89.92%

Mined Dense Corners has a higher classification accuracy than other published methods. This is because of the ability of the technique to select optimal low level features for discrimative classification.

7 Conclusion

This paper has presented a method to efficiently learn informative and descriptive local features of actions performed by humans at multiple scales and temporal speeds. Very coarse corner descriptors are grouped spatially to form an over complete set of feature sets that encode local feature layout. The frequently reoccurring features are then learnt in a weakly-supervised approach where only class labels are required using a data mining algorithm. When tested on the popular KTH dataset, impressive results are obtained which outperform other state-of-the-art approaches while maintaining real-time operation (12fps) in an unoptimised implementation. Although no object segmentation is required during training. The final classifiers can be used to perform activity localisation as well as classification.

Acknowledgments

This work is supported by the EU FP6 Project URUS, and the EU FP7 Project DIPLECS.

References

1. Schuldt, C., Laptev, I., Caputo, B.: Recognizing Human Actions: a Local SVM Approach. In: Proc. of International Conference on Pattern Recognition (ICPR 2004), vol. III, pp. 32–36 (2004)
2. Viola, P., Jones, M.: Rapid Object Detection using a Boosted Cascade of Simple Features. In: Proc. of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2001), vol. I, pp. 511–518 (2001)
3. Ke, Y., Sukthankar, R., Hebert, M.: Efficient Visual Event Detection using Volumetric Features. In: Proc. of IEEE International Conference on Computer Vision (ICCV 2005) (2005)
4. Cooper, H.M., Bowden, R.: Sign Language Recognition Using Boosted Volumetric Features. In: Proc. IAPR Conf. on Machine Vision Applications, pp. 359–362 (2007)
5. Dollar, P., Rabaud, V., Cottrell, G., Belongie, S.: Behavior Recognition via Sparse Spatio-temporal Features. In: ICCCN 2005: Proceedings of the 14th International Conference on Computer Communications and Networks, pp. 65–72 (2005)
6. Laptev, I., Pérez.: Retrieving Actions in Movies. In: Proc. of IEEE International Conference on Computer Vision (ICCV 2007) (2007)
7. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules in large databases. In: VLDB 1994, Proceedings of 20th International Conference on Very Large Data Bases, pp. 487–499 (1994)
8. Quack, T., Ferrari, V., Leibe, B., Gool, L.: Efficient Mining of Frequent and Distinctive Feature Configurations. In: Proc. of IEEE International Conference on Computer Vision (ICCV 2007) (2007)
9. Lazebnik, S., Schmid, C., Ponce, J.: Semi-Local Affine Parts for Object Recognition. In: Proc. of BMVA British Machine Vision Conference (BMVC 2004), vol. II, pp. 959–968 (2004)
10. Sivic, J., Zisserman, A.: Video Data Mining using Configurations of Viewpoint Invariant Regions. In: Proc. of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2004), vol. I, pp. 488–495 (2004)
11. Niebles, J.C., Fei-Fei, L.: A Hierarchical Model of Shape and Appearance for Human Action Classification. In: Proc. of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2007) (2007)
12. Scovanner, P., Ali, S., Shah, M.: A 3-dimensional sift descriptor and its application to action recognition. In: Proc. of MULTIMEDIA 2007, pp. 357–360 (2007)
13. Lowe, D.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision 20, 91–110 (2003)
14. Dalal, N., Triggs, B., Schmid, C.: Human Detection using Oriented Histograms of Flow and Apperance. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 428–441. Springer, Heidelberg (2006)
15. Lucas, B., Kanade, T.: An Iterative Image Registration Technique with an Application to Stereo Vision. In: Proc. of 7th International Joint Conference on Artificial Intelligence (IJCAI), pp. 674–679 (1998)

16. Song, Y., Goncalves, L., Perona, P.: Unsupervised Learning of Human Motion. Transactions on Pattern Analysis and Machine Intelligence 25, 814–827 (2003)
17. Tesic, J., Newsam, S., Manjunath, B.S.: Mining image datasets using perceptual association rules. In: Proc. SIAM International Conference on Data Mining, Workshop on Mining Scientific and Engineering Datasets, pp. 71–77 (2003)
18. Ding, Q., Ding, Q., Perrizo, W.: Association rule mining on remotely sensed images using p-trees. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 66–79 (2002)
19. Chum, O., Philbin, J., Sivic, J., Isard, M., Zisserman, A.: Total Recall: Automatic Query Expansion with a Generative Feature Model for Object Retrieval. In: Proc. IEEE International Conference on Computer Vision (ICCV 2007), pp. 1–8 (2007)
20. Harris, C., Stphens, M.: A Combined Corner and Edge Detector. In: Proc. of Alvey Vision Conference, 189–192 (1988)
21. Fleuret, F., Geman, D.: Coarse to Fine Face Detection. International Journal of Computer Vision 41, 85–107 (2001)
22. Agrawal, R., Imielinski, T., Swami, A.: Mining association rules between sets of items in large databases. In: Proc. of the 1993 ACM SIGMOD International Conference on Management of Data SIGMOD 1993, pp. 207–216 (1993)
23. Nowozin, S., Bakir, G., Tsuda, K.: Discriminative Subsequence Mining for Action Classification. In: Proc. of IEEE International Conference on Computer Vision (ICCV 2007), pp. 1919–1923 (2007)
24. Wong, S.F., Cipolla, R.: Extracting Spatio Temporal Interest Points using Global Information. In: Proc. of IEEE International Conference on Computer Vision (ICCV 2007) (2007)
25. Niebles, J., Wang, H., Fei-Fei, L.: Unsupervised Learning of Human Action Categories using Spatial-Temporal Words. In: Proc. of BMVA British Machine Vision Conference (BMVC 2006), vol. III, pp. 1249–1259 (2006)