

In an embodiment a video action detection system according to claim 12 is provided.

Brief description of the drawing

5 In view of the above, according to a first aspect of the present invention, there is provided a method of automated support vector machine detection of presence of an action of a predetermined type of action from video data, the method comprising the steps of
 - detecting spatiotemporal interest points (STIPs) in the video data;
 10 - extracting descriptors from the video data in spatiotemporal areas at the detected spatiotemporal interest points;
 - assigning the detected spatiotemporal interest points to bins in a feature vector based on the extracted descriptors;
 - computing bin values of the feature vector for each respective bin, the bin
 15 value of each bin comprising sum of contributions associated with descriptors that have been assigned to the bin, with a bin dependent adjustment of a size of the sum and/or the contributions of at least one of the bins dependent on the predetermined type of action;
 - computing a support vector machine (SVM) match score from the feature
 20 vector and reference vectors for the predetermined type of action.

According to a second aspect of the present invention, there is provided a method of automated detection of actions from a plurality of predetermined types of action, comprising executing the method in
accordance with the first aspect for each of the plurality of predetermined
 25 types of action respectively, using, for each respective different predetermined type of action, a respective different bin dependent adjustment of a size of the sum and/or the contributions of at least one of the bins dependent on the predetermined type of action and respective
 different sets of reference vectors.

30 According to a third aspect of the present invention, there is provided a computer readable medium, comprising a program of

instructions for a programmable computer system that, when executed by the programmable computer system, will cause the programmable computer system to execute the method according to any one of the preceding aspects.

According to a fourth aspect of the present invention, there is

provided a video action detection system for detecting presence of an action of a predetermined type of action from video data, the system comprising

- a source of video data or video data input;

- a spatiotemporal interest point (STIP) detector coupled to the source of video data or video data input;

- a feature data extractor coupled to the source of video data or video data input and the STIP detector, configured to extract feature data from video data from the source of video data or video data input at STIPs detected by the STIP detector;

- a descriptor quantizer coupled to the descriptor extractor and configured to assign the extracted descriptor values to bins of a feature vector;

- a computing circuit configured to compute bin values of the feature vector for each respective bin, the bin value of each bin comprising a sum of contributions associated with STIPs that have been assigned to the bin, with a bin dependent adjustment of a size of the sum and/or the

contributions of at least one of the bins dependent on the predetermined type of action; and to compute a support vector machine (SVM) match score from the feature vector and reference vectors for the predetermined type of action.

These and other objects and advantageous aspects will be come apparent from a description of exemplary embodiments, using the following figures

Figure 1 shows an action type detection system

Figure 2 shows a flow chart of video processing

Figure 3 shows a flow-chart of support vectors

determination

Figure 4 shows a flow chart of factor selection

Detailed description of exemplary embodiments

5 Figure 1 shows a video processing system for detecting action types, the system comprising an input video data memory 10, a feature processor 12 with an input coupled to input video data memory 10, a weight memory 14, a multiplier 15 with inputs coupled to outputs of feature processor 12 and weight memory 14, a detector 16 with an input coupled to an output of multiplier 15, a control computer 19, a support vector memory 10
10 17 with an output coupled to detector 16, a detection parameter memory 18 with an output coupled to detector 16. A camera 11 is shown coupled to input video data memory 10. Feature processor 12 comprises a feature detector 120 with an input coupled to input video data memory 10, a feature
15 value extractor 122 with inputs coupled to feature detector 120 and input video data memory 10, a histogram binning (sometimes also called quantization) unit 124 with an input coupled to histogram binning unit 124 and a binning parameter memory 126 coupled to histogram binning unit

Claims

1. A method of automated support vector machine detection of presence of an action of a predetermined type of action from video data, the method comprising the steps of
 - detecting spatiotemporal interest points (STIPs) in the video data;
 - 5 - extracting descriptors from the video data in spatiotemporal areas at the detected spatiotemporal interest points;
 - assigning the detected spatiotemporal interest points to bins in a feature vector based on the extracted descriptors;
 - computing bin values of the feature vector for each respective bin, the bin
 - 10 value of each bin comprising sum of contributions associated with descriptors that have been assigned to the bin, with a bin dependent adjustment of a size of the sum and/or the contributions of at least one of the bins dependent on the predetermined type of action;
 - computing a support vector machine (SVM) match score from the feature
 - 15 vector and reference vectors for the predetermined type of action.
2. A method according to claim 1, comprising using a parameter to control said adjustment of the size of the sum and/or the contributions, by making the size dependent on a value of the control parameter, the method comprising selecting the reference vectors and the value of the parameter,
- 20 the method comprising
 - determining an initial set of reference vectors by means of an automatic training process using training bin values obtained without said adjustment;
 - computing bin scores for individual bins from scores for matches between
 - 25 bin values of bins of the set of reference vectors and counts of STIPs from a set of video segments that have been designated to show the predetermined type of action;

- determining the value of the parameter based on the bin scores for the individual bins, the parameter being set to increase or decrease the bin value of the at least one of the bins relative to the bin values of other ones of the bins according to whether the bin score for the at least one of the bins is larger or smaller than the bin score for the other ones of the bins respectively;

- determining the reference vectors for use in the support vector machine by means of an automatic training process using bin values obtained with said adjustment controlled by the parameter value.

3. A method according to claim 2, wherein the parameter is a weighting factor, the size of the sum and/or the contributions being adjusted by multiplying a count of the STIPs that have been assigned to the bin, and/or contributions to said count by from individual STIPs that have been assigned to the bin, by the weighting factor.

4. A method according to ~~any one of the preceding claims~~claim 1, wherein the weighting factor for the bin is selected dependent on a count value of STIPs in the bin based on a plurality of support vectors for the predetermined type of action, the weighting factor being increased with increasing count value when at least one of the support vectors for the predetermined type of action defines a component value for the bin that is higher than the count value of STIPs in the bin.

5. A method according to claim 2, ~~3 or 4~~, wherein the bin score for each individual bin is computed as a sum of a bin dependent bias value and the score for the match between the bin value and the count of STIPs assigned to the bin, , the method comprising

- determining an overall bias value for the predetermined type of action by means of said automatic training process using training bin values obtained without said adjustment;

- assigning fractions of the overall bias value, for use as the bin dependent bias value, to respective ones of the bins, dependent on said counts for the respective one of bins;

6. A method according to claim 5, wherein the fractions of the overall bias value are assigned to respective ones of the bins in proportion to an average of said counts for the respective one of bins, the average being taken over said set of video segments, or the fractions of the overall bias value are assigned to respective ones of the bins in proportion to a value of said counts for the respective one of bins for one of said set of video segments.

7. A method according to claim ~~5 or 6~~, the method comprising

- computing values of the parameter for a plurality of different ways of assigning fractions of the bias value to respective ones of the bins, including said assignment dependent on said counts for the respective one of bins,
- comparing reliability of the detection of the presence of actions of the predetermined type obtained by detection using respective ones of the parameter values;
- selecting one of the parameter values that yields a highest reliability;
- using the reference vectors by means of an automatic training process using bin values obtained with said adjustment with said adjustment controlled by selected one of the parameter values.

8. A method according to ~~any one of the preceding claims~~claim 1, wherein the bin dependent adjustment comprises multiplying the sum and/or the contributions of at least one of the bins by a weighing factor that depends on the at least one of the bins and the predetermined type of action.

9. A method according to ~~any one of the preceding claims~~claim 1, wherein the SVM match score is computed by computing mutual scores between the feature vector and each the reference vectors for the predetermined type of action, and summing products of the mutual scores with factors for the reference vectors.

10. A method according to claim 9, wherein the mutual scores between the feature vector and the reference vectors are computed for each of the reference vector from a sum over the bins of the smallest of the bin value of the feature vector for the bin and the bin value the reference vector
5 for the bin.
11. A method according to ~~any one of the preceding claims~~claim 1, comprising determining whether a result of said summing the products exceeds a threshold value.
12. A method of automated detection of actions from a plurality of
10 predetermined types of action, comprising executing the method of claim 1 for each of the plurality of predetermined types of action respectively, using, for each respective different predetermined type of action, a respective different bin dependent adjustment of a size of the sum and/or the contributions of at least one of the bins dependent on the predetermined
15 type of action and respective different sets of reference vectors.
13. A computer ~~program product~~readable medium, comprising a program of instructions for a programmable computer system that, when executed by the programmable computer system, will cause the programmable computer system to execute the method according to any one
20 of the preceding claims.
14. A video action detection system for detecting presence of an action of a predetermined type of action from video data, the system comprising
- a source of video data or video data input;
 - a spatiotemporal interest point (STIP) detector coupled to the source of
25 video data or video data input;
 - a feature data extractor coupled to the source of video data or video data input and the STIP detector, configured to extract feature data from video data from the source of video data or video data input at STIPs detected by the STIP detector;
 - 30 - a descriptor quantizer coupled to the descriptor extractor and configured to assign the extracted descriptor values to bins of a feature vector;

- a computing circuit configured to compute bin values of the feature vector for each respective bin, the bin value of each bin comprising a sum of contributions associated with STIPs that have been assigned to the bin, with a bin dependent adjustment of a size of the sum and/or the contributions of at least one of the bins dependent on the predetermined type of action; and to compute a support vector machine (SVM) match score from the feature vector and reference vectors for the predetermined type of action.

15. A video action detection system according to claim 14, comprising
- 10 a training module, the training module being configured to
- determine a set of reference vectors by means of an automatic training process using training bin values obtained without said adjustment;
 - compute bin scores for individual bins from scores for matches between bin values of bins of the further set of reference vectors and counts of
 - 15 spatiotemporal points of interest from a set of video segments that have been designated to show the predetermined type of action;
 - determine the value of the parameter based on the bin scores for the individual bins, the parameter being set to increase or decrease the bin value of the at least one of the bins relative to the bin values of other ones of
 - 20 the bins according to whether the bin score for the at least one of the bins is larger or smaller than the bin score for the other ones of the bins respectively;
 - determine the reference vectors by means of an automatic training process using bin values obtained with said adjustment controlled by the parameter
 - 25 value.