

CLAIMS

What is claimed is:

1. A method for human motion recognition comprising:
decomposing a video sequence into a plurality of atomic actions;
extracting features from each of the plurality of atomic actions, the features comprising at least a motion feature and a shape feature;
with respect to the each of the plurality of atomic actions, convolving histograms of the features to obtain an atomic action histogram; and
performing motion recognition for the video sequence based on a histogram of visual codewords, wherein each bin represents one of the visual codewords and is weighted by an aggregation of distances between the one of the visual codewords with assigned atomic action~~each of the plurality of atomic actions in response to the features.~~
2. The method in accordance with Claim 1, further comprising:
normalizing the histogram of visual codewords such that the bins sum up to unity,~~wherein the step of performing motion recognition for each of the plurality of atomic actions comprises performing motion recognition for each of the plurality of atomic actions by convolving histograms of the features of each of the plurality of atomic actions.~~
3. The method in accordance with Claim 1 or 2, further comprising:
clustering all of the atomic action histograms of the video sequence with a K-means operation to derive cluster centroids as the visual codewords; and

assigning each of the plurality of atomic actions to a nearest one of the visual codewords.

~~34.~~ The method in accordance with any one of Claims 2-1 to 3, wherein the step of extracting features from each of the plurality of atomic actions comprises extracting a set of shape vectors depicting shape flow from each of the plurality of atomic actions.

~~45.~~ The method in accordance with Claim ~~3-4,~~ wherein convolving histograms of the features of each of the plurality of atomic actions comprises deriving a shape descriptor by determining a histogram-of-oriented gradient of the set of shape vectors for each of the plurality of atomic actions.

~~56.~~ The method in accordance with any one of Claims 2-1 to 3, wherein the step of extracting features from each of the plurality of atomic actions comprises extracting a set of motion vectors depicting motion flow from each of the plurality of atomic actions.

~~67.~~ The method in accordance with Claim ~~5-6,~~ wherein convolving histograms of the features of each of the plurality of atomic actions comprises deriving a motion descriptor by determining a histogram-of-oriented optical flow of the set of motion vectors for each of the plurality of atomic actions.

~~78.~~ The method in accordance with any of Claims 2 to 6claims 1 to 7, wherein the features further comprise a pose feature, and wherein the step of

extracting features from each of the plurality of atomic actions comprises extracting a set of pose vectors from each of the plurality of atomic actions, and wherein convolving histograms of the features of each of the plurality of atomic actions comprises deriving a pose descriptor by determining a histogram-of-oriented gradient of the set of shape vectors for each of the plurality of atomic actions.

~~89.~~ The method in accordance with ~~any of the preceding claims~~claim 8, wherein the features further comprise a spatial feature, and wherein the step of extracting features from each of the plurality of atomic actions comprises deriving each of a ~~the~~ set of shape, motion or pose vectors for each of two or more regions of a bounding box within each of the plurality of atomic actions, the bounding box in each of the plurality of atomic actions configured to include all of a subject pictured in the one of the plurality of atomic actions, and wherein the step of ~~performing motion recognition for each of the plurality of atomic actions~~ comprises convolving histograms of the features of each of the plurality of atomic actions comprising convolving each of the shape, motion or pose descriptors to generate a resultant histogram.

~~910.~~ The method in accordance with any of Claims ~~2-1~~ to ~~8-9~~ wherein convolving histograms of the features of each of the plurality of atomic actions comprises normalization of the histograms of each of the plurality of atomic actions to sum up to unity.

~~10.~~ The method in accordance with any of the preceding claims further comprising K-means clustering of all of the atomic actions to generate a distance-weighted bag-of-automatic-actions model of the video sequence.

11. A system for human motion recognition comprising:

a video sequence decomposer for decomposing a video sequence into a plurality of atomic actions;

a feature extractor for extracting features from each of the plurality of atomic actions, the features comprising at least a motion feature and a shape feature; and

a motion recognition module for performing motion recognition ~~for~~, wherein the motion recognition module is configured to:

with respect to the each of the plurality of atomic actions, convolve histograms of the features to obtain an atomic action histogram; and

perform motion recognition for the video sequence based on a histogram of visual codewords, wherein each bin represents one of the visual codewords and is weighted by an aggregation of distances between the one of the visual codewords with assigned atomic action~~each of the plurality of atomic actions~~in response to the features.

12. The system in accordance with Claim 11 wherein the motion recognition module is further configured to normalize the histogram of visual codewords such that the bins sum up to unity~~performs motion recognition for each of the plurality of atomic actions by convolving histograms of the features of each of the plurality of atomic actions.~~

13. The system in accordance with Claim 11 or 12, wherein the motion recognition module is further configured to:

cluster all of the atomic action histograms of the video sequence with a K-means operation to derive cluster centroids as the visual codewords; and

assign each of the plurality of atomic actions to a nearest one of the visual codewords.

~~13~~14. The system in accordance with any one of Claims 11 to ~~12~~ 13, wherein the feature extractor extracts a set of shape vectors depicting shape flow from each of the plurality of atomic actions.

~~14~~15. The system in accordance with Claim ~~13~~ 14, wherein the motion recognition module convolves histograms of the shape features of each of the plurality of atomic actions by deriving a shape descriptor by determining a histogram-of-oriented gradient of the set of shape vectors for each of the plurality of atomic actions.

~~15~~16. The system in accordance with any one of Claims ~~12~~ 11 to 13, wherein the feature extractor extracts a set of motion vectors depicting motion flow from each of the plurality of atomic actions.

~~16~~17. The system in accordance with Claim ~~15~~ 16, wherein the motion recognition module convolves histograms of the motion features of each of the plurality of atomic actions by deriving a motion descriptor by determining a histogram-of-oriented optical flow of the set of motion vectors for each of the plurality of atomic actions.

~~17~~18. The system in accordance with any of Claims ~~12~~ 11 to ~~16~~ 17, wherein the features further comprise a pose feature, and wherein the feature extractor further extracts a set of pose vectors from each of the plurality of atomic actions, and wherein the motion recognition module convolves histograms of the pose features of each of

the plurality of atomic actions by deriving a pose descriptor by determining a histogram-of-oriented gradient of the set of pose vectors for each of the plurality of atomic actions.

~~18~~19. The system in accordance with ~~any of Claims 12 to 17~~18, wherein the features further comprise a spatial feature, and wherein the feature extractor derives each of ~~a the~~ set of shape, motion or pose vectors for each of two or more regions of a bounding box within each of the plurality of atomic actions, the bounding box in each of the plurality of atomic actions configured to include all of a subject pictured in the one of the plurality of atomic actions, and wherein the motion recognition module convolves histograms of each of the shape, motion or pose descriptors to generate a ~~resultant~~the atomic action histogram.

~~19~~20. The system in accordance with any of Claims ~~12-11~~ to ~~18-19~~, wherein the motion recognition module normalizes the histograms of each of the plurality of atomic actions to sum up to unity.

~~20. The system in accordance with any of Claims 11 to 19 wherein the motion recognition module further comprises a bag-of-words model module for K-means clustering of all of the atomic actions to generate a distance weighted bag-of-automatic-actions model of the video sequence.~~