Patent Specification

confidential

Title of the Invention

Methods and Apparatus for Person Identification and Person Re-Identification with full or partially available parts in a structured feature set

Inventors

Shen Shengmei, Wang Hao, Chen Yunpeng, Chang Suning, Feng Jiashi

Pensees Pts Ltd

3 Fusionopolis Way, #14-23/24 Symbiosis, Singapore 138633

Field of the Art

The present invention relates to the field of Person Identification and Person Re-Identification from same or different camera views. Specifically, the present invention makes use of multiple features of a person including face features, ear features, features from head and other body parts, from coarse to fine-grained feature, from global to local features to improve the robustness and efficacy of the person recognition or person re-identification no matter how a person's view to be seen and captured from a camera, or how a person's part is occluded from the real captured scene.

Not only can it be applied to generic Person Re-Identification, but also applied to a finer level which consists of face or body parts (e.g. limbs, torsos, heads). This technique can be used to describe a person in a structure way. The invention makes person search easier with the said structured multiple features of a person even in the case of occlusion case where certain part of person body is missing. Furthermore unique features of a person like face features, ear features can be highly considered with higher weights for re-identifying as the same person whenever such features are available.

Background Art

Person Identification and Person Re-Identification technology has made great impact on the real application with the advancement of deep learning technologies. With the help of Deep Neural Networks, many real challenges have been addressed, including variations of illumination, pose and occlusion.

However the performance is still very poor when applying in real application due to unconstrained camera captured condition where a person's view is changed drastically and part of the said person is occluded, a global feature of a person no longer can handle the problem well.

On the other hand, person re-identification approach only considers general person's feature without considering part-based feature set and their structured combination especially some unique feature set like face or ear to give strong clue of the person re-identification.

Summary of the Invention

Problems to be solved

The aforementioned prior art has the following problems:

Existing Person Identification and Person Re-Identification methods fail to match images from the same person but with different views;

Existing Person Identification and Person Re-Identification methods fail to match images from the same person but with occlusion of certain part;

Existing Person Identification and Person Re-Identification methods did not utilize other unique features like face or ear to enhance the matching performance;

Existing Person Identification and Person Re-Identification methods cannot tell which part of body between the query and the gallery images are similar, degrading so much for the performance of the matching and re-identifying of a person for many cases.

Means of Solving the Problems

The proposed invention utilizes a multi-channel pipeline to extract features of different body parts including face, ear and other parts of the said person.

The proposed invention stores a structured feature sets which are corresponding to person body part including face, ear and other parts.

Effects of the Invention

This invention is useful because it makes the Neural Network pay attention not only to global features but also local features, which could boost the accuracy of the feature extractor. Both the local feature and the global feature are applicable to take part in the person searching process. The benefit is that now the searching is more flexible and the feature extractor becomes interpretable because it can tell us based on which part does it make its decision.

The present invention provides a structured and fine-grained Person Identification and Person Re-Identification method including face and ear part. This method makes person searching easier even with partial part of the said person or partial information of the structured feature sets. Also, it increases the recall rate of relevant persons, especially when only partial of a person information is available.

Best mode for carrying out the invention

1. Spatial feature grouping

Assuming we have a feature-map derived from a Convolutional Neural Network, we split the feature map vertically into number of sub-feature maps, here we use 6 as an example. Then we can have $2^6 - 2 = 62$ combinations. Each combination's sub-feature maps are concatenated vertically to form a new feature map. Subsequently, each new feature map goes through the Global Average Pooling layer and becomes a feature vector. All feature vectors are compressed to lower dimensions and go to two branches. One branch uses all compressed feature vectors to calculate the Triplet loss. The other branch calculates an ID for each of the compressed feature vector.

2. Part Segmentation

Say we have six body parts. Then we can estimate each part's spatial area by image segmentation. Then, for each part, we average the initial feature map over the part's area instead of GAP as conventional methods do. Thus, we receive six part's feature vectors. They are sent to a Graph Convolutional Network, which learns relationships among the parts, to form a global feature vector. Then we calculate a Triplet loss and an ID loss to guide the training.

3. Feature Pyramid reconstruction

The calculation is conducted in real-time. First, we down-sample the feature map extracted by our Deep Neural Network recursively to build the feature pyramid. Next, we try to find a linear combination of the gallery image's feature pyramid so that the feature pyramid error between the query and the linear combination is minimized. Finally, we regard the similarity of the feature pyramids between the gallery linear combination and the query as the estimated similarity that is used in the comparison process.

Figure 1.1 shows the general system data processing flow, here human and person are considered as the same meaning.

Input image including from various sensors goes to a Human detection engine in the processing block of 100. Once a human is detected regardless it is full part or partial parts of human and regardless what kind of pose of the said human is, Human Parsing in the block of 101 is performed to segment a human into number of parts.

The human parsing information is passed to Person Identification or Person Re-identification engine in the block of 102, to perform global Person feature extraction and local person feature extraction, to name as Global Feature1 and Local Feature sub-set_a.

The human parsing information is also passed to Face and head detection engine in the block of 103. If Face is detected in the block of 103, Face parsing will output face local regions in the block of 104 then going through face identification or face re-identification engine in the block of 105, to output Global Feature2 and Local Feature sub-set_b, or if head part is detected in the block of 103, ear detection in the block of 106 is performed following by ear identification or ear re-identification engine in the block of 107.

The human parsing information is passed to Person Attributes Estimation in the block of 108 to estimate the person attribute feature sets.

Figure 1.2 shows the said Structured Feature which is formed from the above mentioned Global Feature1, 2, … and Local Feature sub-set_a, b,…. In the case of some feature set is not available either in gallery image or query image, the similarity comparison for this part in the decision will be excluded so to produce more accurate decision with the corresponding parts which are available in both gallery image or query image. Weights for different feature set will be considered depending on the importance and confidence level of each feature set.

Figure 2 shows embodiment 1 where the data flow when conducting the aforementioned spatial sub-feature map combination. For example, we split the feature map into vertically distributed number of sub-feature maps, here 6 will be used for explanation. Then we take all possible combination of the 6 sub-feature maps except the full set and the empty set. In this manner, we can have $2^6 - 2 = 62$ combinations. Then we assemble all sub-feature maps within each combination and perform a global pooling to get a feature vector for each spatial combination. Each spatial combination branch is followed by a 1×1 Convolutional layer, a Batch Normalization layer and a ReLU activation layer. All these derived feature vectors are concatenated into a long feature vector which is supervised by a Triplet Loss. Meanwhile, each sub-feature vector is compressed (in accordance with the number of ID's) by a Full Connection layer and then supervised by an ID Loss. The next part of the figure shows a way to utilize both the global and the local features. The distance could be either Euclidian or Cosine distance.

Figure 3 shows the Reconstruction and inference process mentioned stated in Figure 2, which is used in feature comparison phase. Assuming we have a query image and a gallery image. First, we extract each image's feature map using the Deep Neural Network aforementioned. Then we down-sample the feature maps recursively to build a spatial feature pyramid for both the query and the gallery. Then optimize a weight matrix so that the query feature pyramid is closer (measured by Euclidean distance) to a linear combination of the gallery feature pyramid. Finally, the closest distance we can achieve is treated as the distance in the comparison process.

Figure 4 shows another embodiment where the data flow of Networks using image segmentation guided pooling features. First, we have a common feature map that could be calculated by any Convolutional Neural Network. Then we place a 1×1 Convolutional layer

to learn the whole-body segmentation. At the same time, the segmentation is regarded as a mask that helps filtering out background information. The same manner applies to other human body parts as illustrated in the figure. All the feature maps passed through the filtering process are polled globally to derive the relevant feature vectors. For the body part feature vectors, to model their relations, we introduce a GCN:Graph Convolution Network layer to merge the parts' feature vectors as a single one. Then the feature vectors are supervised by a Triplet loss and an ID loss.

Figure 5 shows the structure of the previously introduced Graph Convolutional Network.

To start with, we stack the parts' feature vectors to form a feature map representing human body as a graph with each row as a body part's feature vector. Then we perform the same dimensional Convolution for each column to model the body graph's diffusional characteristics (i.e. Laplace Matrix). Afterwards, we transpose the body feature map and perform another one-dimensional Convolution on each part's feature vector. At last, we restore the shape of the graph feature by transpose it back. At this point, a whole Graph Convolutional Network layer's step is explained. We apply this same step by K times so that information is sufficiently passed among parts and within feature vectors.

Industry Applications

The Person Identification and Person Re-Identification technology can be used as a video infrastructure to aggregate other information for many applications. With the help of this invention, person's profile could be established to form person big database which has more information than face profile database.

In addition, Person Identification and Person Re-Identification technology can also directly be used to search for persons of interest. Especially in the forensic, searching for suspects by outlooks demands for a huge amount of time and effort when it is done manually. With this technology, it is now possible to find a suspect within seconds among a big number of camera channels or in a huge dataset.

Brief description of the drawings

Figure 1.1 shows the general system data processing flow;

Figure 1.2 shows the Structured Feature;

Figure 2 shows one of the combinations of feature sets;

Figure 3 shows another feature pyramid reconstruction and comparison;

Figure 4 shows Body Parts Segmentation polling;

Figure 5 shows Graph Convolutional Network;

What is claimed is:

1. Methods and Apparatus for Person Identification and Person Re-Identification with full or partially available parts in a structured feature set, comprising of:

   Detecting and tracking person in a video or a sequence of images;

   Parsing and segmenting each of the said person into several parts for further processing;

   Extracting each of the said person's global features and its local feature set for each of the parts of the said person;

   Detecting face and/or head region from the said segmented parts of the said person;

   Passing the said detected face to face recognition engine for further processing.

2. Methods and Apparatus for Person Identification and Person Re-Identification with full or partially available parts in a structured feature set, where Passing the said detected face to face recognition engine in the claim 1, further comprising of:

   Parsing and segmenting the said face into several parts;

   Extracting face global feature and its local feature set for each of the said parts of the said face.

3. Methods and Apparatus for Person Identification and Person Re-Identification with full or partially available parts in a structured feature set, where Detecting face and/or head region from the said segmented parts of the said person in the claim 1, further comprising of:

   Extracting ear feature set if ear is detected.

4. Methods and Apparatus for Person Identification and Person Re-Identification with full or partially available parts in a structured feature set, where Parsing and segmenting each of the said person into several parts for further processing in the claim 1, further comprising of:

   Estimating the said person's attributes and extracting the corresponding feature set.

5. Methods and Apparatus for Person Identification and Person Re-Identification with full or partially available parts in a structured feature set, where Extracting each of the said person's global features and its local feature set for each of the parts of the said person in the claim 1, further comprising of:

   Combining local feature sets from different parts of the said person in a pyramid structure.

6. Methods and Apparatus for Person Identification and Person Re-Identification with full or partially available parts in a structured feature set, comprising of:

   Aggregating and combining various feature sets from various parts of the said person including their global feature set and local feature set;

   Forming a structured feature configuration with signaling of available or unavailable for each feature set;

   Signaling the said feature set as unavailable if the corresponding part of the said person is not available, otherwise Signaling the said feature set as available if the corresponding part of the said person is available;

   Signaling the said feature set as unavailable if the corresponding part of the said face is not available, otherwise Signaling the said feature set as available if the corresponding part of the said face is available;

   Calculating separate similarity scores only for each of the said available feature sets of the sad person if the corresponding parts of the said person in both gallery and query are available;

   Combining all the calculated similarity scores into total score;

   Comparing with a threshold and determine if the same person is identified.

7. Methods and Apparatus for Person Identification and Person Re-Identification with full or partially available parts in a structured feature set, where Forming a structured feature configuration in the claim 6, further comprising of:

   Constructing the mapping between feature sets and the corresponding parts using graph convolutional network.

8. Methods and Apparatus for Person Identification and Person Re-Identification with full or partially available parts in a structured feature set, where Combining all the calculated similarity scores into total score in the claim 6, further comprising of:

   Assigning different weights to different sets of features according to their importance and confidence level.