ECOC and Graph Cuts
Segmentation of Human Limbs

Author: Daniel Sánchez
Advisors: Sergio Escalera and Miguel Ángel Bautista
1. Motivation
2. HuPBA 8k+ dataset
3. Proposal
4. Results
5. Conclusions
• User Detection/Segmentation

• Applications: medicine, photography, sign language...
- **Tasks:** Multi-limb human pose detection, segmentation, action /gesture recognition.

- 9 actors, 14 limbs categories, 11 gesture categories (isolated and collaborative actions).

- More than 8,000 frames, 120,000 manual labeled limbs at pixel precision.


- Action categories: Wave, Point, Clap, Crouch, Jump, Walk, Run, Shake Hands, Hug, Kiss and Fight.
HuPBA 8k+ dataset

- Compare with some publicly available datasets. These public datasets are chosen taking into account the variability of limbs and gestures/actions.

<table>
<thead>
<tr>
<th></th>
<th>HuPBA PARSE</th>
<th>BUFFY</th>
<th>UIUC people</th>
<th>LEEDS SPORTS</th>
<th>HW</th>
<th>MMGR13</th>
<th>H.Actions</th>
<th>Pascal VOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labeling at pixel precision</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Number of limbs</td>
<td>14</td>
<td>10</td>
<td>6</td>
<td>14</td>
<td>14</td>
<td>-</td>
<td>16</td>
<td>-</td>
</tr>
<tr>
<td>Number of labeled limbs</td>
<td>124 761</td>
<td>3 050</td>
<td>4 488</td>
<td>18 186</td>
<td>28 000</td>
<td>-</td>
<td>27 532 800</td>
<td>-</td>
</tr>
<tr>
<td>Number of frames</td>
<td>8 234</td>
<td>305</td>
<td>748</td>
<td>1 299</td>
<td>2 000</td>
<td>-</td>
<td>1 720 800</td>
<td>-</td>
</tr>
<tr>
<td>Full body</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>-</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Limb annotation</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Gesture annotation</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Number of gestures</td>
<td>11</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>8</td>
<td>20</td>
<td>6</td>
</tr>
<tr>
<td>Number of gesture samples</td>
<td>235</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>430</td>
<td>13 858</td>
<td>600</td>
</tr>
</tbody>
</table>
Proposal

Results

Conclusions

Stage 1: Binary Segmentation

(a) Input
   → Tree-structure learning of human limbs
   - Haar-like features
   - Cascade of AdaBoost

(b) ECOC multi-limb detection
   - Loss-Weighted decoding

(c) Body-like probability map

(d) Binary GrabCut optimization for foreground mask extraction

Stage 2: Multi-limb Segmentation

(e) Person/Background segmentation input

(f) Tree-structure body part learning without background

(g) ECOC multi-limb detection
   - HOG features
   - SVM+RBF Kernel
   - Loss-Weighted decoding

(h) Limb-like probability map definition

(i) Alpha-beta swap Graph Cuts multi-limb segmentation
Body part learning using cascade of classifiers

- Body parts rotational invariant by computing dominant orientation.

- Haar-like features describes those body parts.

- Adaboost as the base classifier in the cascade architecture.

---


Define the groups of limbs to be learnt by each individual cascade.

Proposal

Stage 1: Binary Segmentation

(a) Input

- Tree-structure learning of human limbs
  - Haar-like features
  - Cascade of AdaBoost

(b) ECOC multi-limb detection
  - Loss-Weighted decoding

(c) Body-like probability map

(d) Binary GrabCut optimization for foreground mask extraction

Stage 2: Multi-limb Segmentation

(e) Person/Background segmentation input

(f) Tree-structure body part learning without background

(g) ECOC multi-limb detection
  - HOG features
  - SVM-RBF Kernel
  - Loss-Weighted decoding

(h) Limb-like probability map definition

(i) Alpha-beta swap Graph Cuts multi-limb segmentation
In classification tasks, the goal is to classify an object among a certain number of possible categories.

This framework is composed of two different steps:
- **Coding**: Decompose a given $N$-class problem into a set of $n$ binary problems.
- **Decoding**: Given a test sample $s$, determine its category.

![Diagram](https://via.placeholder.com/150)
Proposal

- At the decoding step a new sample $s$ is classified by comparing the binary responses to the rows of $M$ by means of a decoding measure $\delta$.

- Different types of decoding based on the distance used (i.e. Hamming, Euclidean, etc.)

$$\arg\min_i \delta(x^s, y^i)$$
We propose to use a predefined coding matrix in which each dichotomy is obtained from the body part tree-structure.

Then, each cascade will give us its prediction and decoding ECOC step will be applied.

- Loss-weighted decoding using cascade of classifier weights (takes into account classifiers performance)

In order to classify a new sample we apply a sliding window over the image:
Proposal

Results

Conclusions

Stage 1: Binary Segmentation

Stage 2: Multi-limb Segmentation

(a) Input

(b) Tree-structure learning of human limbs
- Haar-like features
- Cascade of AdaBoost

(c) ECOC multi-limb detection
- Loss-Weighted decoding

(d) Body-like probability map

(e) Person/Background segmentation input

(f) Tree-structure body part learning without background

(g) ECOC multi-limb detection
- HOG features
- SVM-RBF Kernel
- Loss-Weighted decoding

(h) Limb-like probability map definition

(i) Alpha-beta swap Graph Cuts multi-limb segmentation
A body-like probability map \( P^{bl} \in [0, 1]^{l \times w} \) is built.
Proposal

○ Image Segmentation == Image labeling!

Results

○ Graph Cuts (Energy minimization)

![Graph Cuts](image)

- User interaction by superimposed user input, background brush and so on.

- Binary segmentation by means of background and foreground segmentation.
  - Background
  - Foreground

Human Pose Recovery and Behavior Analysis Group

Proposal

Results

Conclusions

Stage 1: Binary Segmentation

(a) Input
(b) Tree-structure learning of human limbs
   • Haar-like features
   • Cascade of AdaBoost
(c) ECOC multi-limb detection
   • Loss-Weighted decoding
(d) Body-like probability map
   • Binary GrabCut optimization for foreground mask extraction

Stage 2: Multi-limb Segmentation

(e) Person/Background segmentation input
(f) Tree-structure body part learning without background
(g) ECOC multi-limb detection
   • HOG features
   • SVM-RBF Kernel
   • Loss-Weighted decoding
(h) Limb-like probability map definition
(i) Alpha-beta swap Graph Cuts multi-limb segmentation
Define the groups of limbs to be learnt by each individual cascade without background.

Proposal

Results

Conclusions

Stage 1: Binary Segmentation

(a) Input

(b) Tree-structure learning of human limbs
   - Haar-like features
   - Cascade of AdaBoost

(c) ECOC multi-limb detection
   - Loss-Weighted decoding

(d) Body-like probability map

(e) Binary GrabCut optimization for foreground mask extraction

Stage 2: Multi-limb Segmentation

(e) Person/Background segmentation input

(f) Tree-structure body part learning without background

(g) ECOC multi-limb detection
   - HOG features
   - SVM+RBF Kernel
   - Loss-Weighted decoding

(h) Limb-like probability map definition

(i) Alpha-beta swap

Graph Cuts multi-limb segmentation
## Proposal

- Body parts rotational invariant by computing dominant orientation.

## Results

- HOG features describes those body parts.

## Conclusions

- SVM classifiers with Generalized Gaussian RBF Kernel based on Chi-squared distance.

---


We propose to use a predefined coding matrix in which each dichotomy is obtained from the body part tree-structure without background.

Then, each cascade will give us its prediction and decoding ECOC step will be applied.

- Loss-weighted decoding using cascade of classifier weights (takes into account classifiers performance)

In order to classify a new sample we apply a sliding window over the image considering the binary mask:
Proposal

Results

Conclusions

Stage 1: Binary Segmentation

(a) Input
(b) Tree-structure learning of human limbs
   - Haar-like features
   - Cascade of AdaBoost
(c) ECOC multi-limb detection
   - Loss-Weighted decoding
(d) Body-like probability map
   - Binary GrabCut optimization for foreground mask extraction

Stage 2: Multi-limb Segmentation

(e) Person/Background segmentation input
(f) Tree-structure body part learning without background
(g) ECOC multi-limb detection
   - HOG features
   - SVM+RBF Kernel
   - Loss-Weighted decoding
(h) Limb-like probability map definition
(i) Alpha-beta swap Graph Cuts multi-limb segmentation
Limb-like probability maps are build

- Haar-like based on AdaBoost gave us an accurate and efficient initialization of human regions for binary user segmentation.
- HOG-SVM is applied in a reduced region of the image, providing better estimates of human limb locations.
Proposal

Stage 1: Binary Segmentation

(a) Input
   → Tree-structure learning of human limbs
      • Haar-like features
      • Cascade of AdaBoost

(b) ECOC multi-limb detection
   • Loss-Weighted decoding

(c) Body-like probability map

(d) Binary GrabCut optimization for foreground mask extraction

Stage 2: Multi-limb Segmentation

(e) Person/Background segmentation input

(f) Tree-structure body part learning without background

(g) ECOC multi-limb detection
   • HOG features
   • SVM-RBF Kernel
   • Loss-Weighted decoding

(h) Limb-like probability map definition

(i) Alpha-beta swap Graph Cuts multi-limb segmentation

Results

Conclusions
o Multi-label segmentation problem $\rightarrow$ Graph Cuts alpha-beta swap

o Segmentation by combining all pair labels \[ (\alpha_q, \alpha_m), \{m, q\} \in \{1, 2, \ldots, 6\} \]

o A predefined pair-wise cost function $\Omega(c_q, c_m)$ penalizes relations between labels taking into account the natural constraints of the human limbs.
DATA

- HuPBA 8k+: Combining by symmetry we obtain 6 limbs categories.
- Limb regions are scale and rotational invariant for training.
- Resize all limb sample to a 32x32 pixels region for computational purposes.
- Limbs categories and samples: head (9,000), torso (9,000), left/right arms (14,000), left/right forearms (15,200), left/right thighs (16,900), left/right legs (17,200).
- Action categories and samples: Wave (10), Point (13), Clap (15), Crouch (10), Jump (20), Walk (72), Run (17), Shake Hands (), Hug (18), Kiss (18), Fight (18).
RESULTS

FIRST STAGE (Binary Segmentation)
- Haar-like features + AdaBoost: forced a 0.9 false positive rate and maximum of 0.4 false alarm rate during 8 stages.
- Test: Sliding window approach with an initial patch size of 32x32 pixels up to 60x60 pixels.
- Use of Graph Cut for binary segmentation tuned via cross-validation.

SECOND STAGE (Multi-limb Segmentation)
- HOG descriptor: 32x32 window size, 16x16 block size, 8x8 block stride, 8x8 cell size and 8 for number of bins.
- SVMs with a Generalized Gaussian RBF kernel based on Chi-squared. The parameters of the kernel were tuned via cross-validation.
- Model selection was done via a leave-one-sequence-out cross-validation.
- Multi-limb segmentation: Alpha-beta GraphCut procedure, we set a 8x8 neighboring grid and tuned the \( \lambda \) parameter using cross-validation.
**GESTURE RECOGNITION**
- Feature vector of a frame: concatenation of the 6 limb-like probability maps, resizing each one of them to a 40x20 pixels region and vectorizing that region. Obtaining a final vector of $d = 40 \times 20 \times 6 = 4800$ dimensions, which is then reduced to $d = 150$ dimensions using a Random Projection.

- Cost-threshold and the action/gesture model for both DTW experiments was obtained by cross-validation on training data, using a leave-one-sequence-out procedure.

- Each HMM and its corresponding probability-threshold was obtained by cross-validation on training data, using a leave-one-sequence-out procedure.

**EVALUATION:**
- Jaccard Index overlapping, $J = \frac{A \cap B}{A \cup B}$

- Do not care value
We compare three methods:


<table>
<thead>
<tr>
<th>Method</th>
<th>Precision (± Standard Deviation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P.Detector + GbCut*</td>
<td>49.60 ± 20.45</td>
</tr>
<tr>
<td>C.Class + GbCut**</td>
<td>58.26 ± 17.31</td>
</tr>
<tr>
<td>C.Class + ECOC + GbCut</td>
<td>61.79 ± 14.02</td>
</tr>
</tbody>
</table>

* (our proposal)
We compare three methods:
- ECOC + GraphCut (our proposal)
- Flexible-mixture-parts (FMP)*
- Iterative Parsing Process (IPP)**

Results: Multi-limb Segmentation

Proposal

Results

Conclusions

**Head limb class**

- **Torso limb class**

- **Arms limb class**

- **Forearms limb class**

- **Thighs limb class**

- **Legs limb class**

Mean Jaccard Index

- Ours
- IPP
- FMP

Number of Do not care frames
Results: Action Recognition

Proposal

Results

Conclusions

Wave action class

Point action class

Clap action class

Crouch action class

Jump action class

Walk action class

Run action class

Shake hands action class

Hug action class

Kiss action class

Fight action class

Mean Jaccard Index

DTW Random

DTW Mean

HMM
We introduce the **HuPBA 8k+ dataset**, the largest RGB labeled dataset of human limbs, with more than 120000 manually annotated limbs. The data set also includes frame-level annotation for 11 action/gesture categories.

We propose a **two stage approach based on ECOC and Graph Cuts** for the segmentation of human limbs in RGB images.

The proposed method is compared with state-of-the-art methods for human pose estimation obtaining very satisfying results.

We provide with a baseline for Action Recognition in the novel dataset.


Selected from IBPRIA and extended version submitted to Neurocomputing journal.

The novel data set is currently used for the ChaLearn posture-gesture recognition challenge and workshop at European Conference on Computer Vision 2014 by the HuPBA group of the University of Barcelona.
ECOC and Graph Cuts
Segmentation of Human Limbs

Thank you!

Author: Daniel Sánchez
Advisors: Sergio Escalera and Miguel Ángel Bautista