An Efficient Branch-and-Bound Algorithm for Optimal Human Pose Estimation

Min Sun Murali Telaprolu Honglak Lee Silvio Savarese
Dept. of Electrical Engineering and Computer Science, University of Michigan at Ann Arbor, USA

1. Overview

GOAL: Propose an efficient and exact inference algorithm based on branch-and-bound (BB) to solve the human pose estimation problem on loopy graphical models

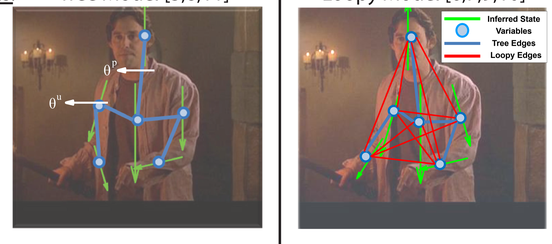
Motivation:

- Cast human pose estimation problem as MAP-MRFs inference problem
- Solving MAP inference on general MRFs is challenging

Given the model $f(\mathrm{h};\theta)=\sum_{i\in N}\theta_i^u(h_i)+\sum_{ij\in\varepsilon}\theta_{ij}^p(h_i,h_j)$,

search for the best body parts locations $\mathrm{h}^{\mathrm{MAP}}=\max_{\mathrm{h}} f(\mathrm{h};\theta)$.

Models: Tree Model [3,8,11] | Loopy Model [6,7,9,10]



Pros: ● efficient inference by dynamic programming ($O(H^2)$).
Cons: ● approximated model; ● common misclassification errors.

Pros: ● more interactions between parts.
Cons: ● exact inference NP-hard; ● require approximated inference [1] or reduced state space.

Contributions:

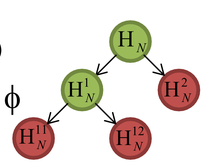
- Solve loopy models with large # of part hypotheses exactly.
- Efficiency is provably guaranteed when # of nodes is small.
- Up to 74 times faster than competing techniques [2] !
- Enable more accurate results than non-loopy models [3] !
- **Source code available online (http://www.eecs.umich.edu/vision/BBproj.html).**

Key Intuitions:

- Flexible bound by relaxing the loopy model into a mixture of star models.
- Novel data structure (BMT) and an efficient search routine (OBMS) to significantly reduce the time complexity for calculating the bound in each branch of the BB search from $O(H^2)$ to $O(Q\log_2 H)$ (Q<<H).

2. Branch-and-Bound Basic

- Bound: $(\mathrm{UB}(\mathrm{H}_N),\mathrm{LB}(\mathrm{H}_N))$ of $f(\mathrm{h}^{\mathrm{MAP}};\theta)$
- Branching: $\mathrm{H}_N=\mathrm{H}_N^1\cup\mathrm{H}_N^2;\mathrm{H}_N^1\cap\mathrm{H}_N^2=\phi$
- Stopping criteria: $\mathrm{UB}(\mathrm{H}_N)=\mathrm{LB}(\mathrm{H}_N)$

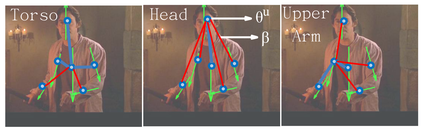


3. Flexible Bound

Mixture of Star Models:

$f^R(\mathrm{h};\theta^u,\beta)=\sum_{i\in N}(\theta_i^u(h_i)+\sum_{j\in N_i}\max_{\hat{h}_j\in\mathrm{H}_j}\beta_{ji}(\hat{h}_j,h_i))$

s.t. $\beta_{ji}(h_j,h_i)+\beta_{ij}(h_i,h_j)=\theta_{ij}^p(h_i,h_j)$



Bounds: (Time complexity $O(H^2)$)

$\mathrm{UB}(\mathrm{H}_N;\theta^u,\beta)=\max_{\mathrm{h}\in\mathrm{H}_N} f^R(\mathrm{h};\theta^u,\beta)\ \geq f(\mathrm{h}^{\mathrm{MAP}};\theta)$

$\mathrm{LB}(\mathrm{H}_N;\theta^u,\beta)=f(\mathrm{h}^*(\mathrm{H}_N);\theta)\leq f(\mathrm{h}^{\mathrm{MAP}};\theta)$ (Constant time)

Tightness of the Bound:

Primal-Dual-Gap (PDG) $=\mathrm{UB}(\mathrm{H}_N;\theta^u,\beta)-\mathrm{LB}(\mathrm{H}_N;\theta^u,\beta)$

- β controls the tightness of the bound.
- PDG reduced by solving β using Message Passing [1].

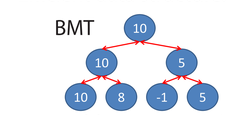
4. Efficient Bound

Naive Bound: time complexity $O(H^2)$; H is #states.

$\mathrm{UB}(\mathrm{H}_N;\theta^u,\beta)=\max_{h_i\in\mathrm{H}_i}\sum_{i\in N}(\theta_i^u(h_i)+\sum_{j\in N_i}\max_{\hat{h}_j\in\mathrm{H}_j}\beta_{ji}(\hat{h}_j,h_i))$

Branch-Max-Tree (BMT): Similar to [5]

- 1D array: $\max_{h_j\in\mathrm{H}_j}\mathrm{A}^{h_i}[h_j]=\max_{h_j\in\mathrm{H}_j}\beta_{ji}(h_j,h_i)$
- Efficient data structure: given A =[10 8 -1 5].



Max(A[0..3]) = 10

Max(A[0,1]) = 10

Max(A[2,3]) = 5

Building time: O(H)

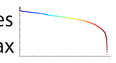
Query time: O(1)

- Time complexity: O(H).

Opportunity Branch-Max-Search (OBMS)

- Re-define $\mathrm{UB}(\mathrm{H}_N;\theta^u,\beta)=\sum_{i\in N}\max_{\mathrm{h}\in\mathrm{H}_N}\lambda_i(h_i;\mathrm{H}_N)$;
- 1D array: $\max_{h_i\in\mathrm{H}_i}\mathrm{A}[h_i]=\max_{h_i\in\mathrm{H}_i}\lambda_i(h_i;\mathrm{H}_N)$

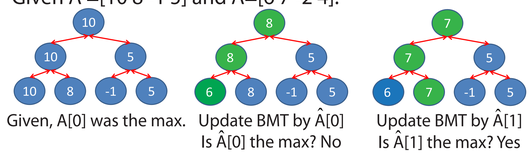
1) only a few states competing for max



2)bound for $\hat{\mathrm{H}}_N\subset\mathrm{H}_N$
$\mathrm{A}[h_i]\geq\hat{\mathrm{A}}[h_i]=\lambda_i(h_i;\hat{\mathrm{H}}_N)$

- Opportunity Search by updating BMT : $\max_{h_i\in\hat{\mathrm{H}}_i}\hat{\mathrm{A}}[h_i]$

Given A =[10 8 -1 5] and Â=[6 7 -2 4].



Given, A[0] was the max.

Update BMT by Â[0] Is Â[0] the max? No

Update BMT by Â[1] Is Â[1] the max? Yes

- Time complexity: $O(Q\log_2 H)$, Q<<H.

5. Branching Strategy

Guided Variable Selection (GVS): Split the selected variable, e.g., human pose: select a specific body part.

- Re-define Upper Bound

$\mathrm{UB}(\mathrm{H}_N;\theta^u,\beta)=\max_{\mathrm{h}\in\mathrm{H}_N} f^R(\mathrm{h};\theta^u,\beta)=\sum_{i\in N}\lambda_i(h_i^*)$;

$\lambda_i(h_i)=\theta_i^u(h_i)+\sum_{j\in N_i}\max_{\hat{h}_j\in\mathrm{H}_j}\beta_{ji}(\hat{h}_j,h_i)$

- Re-define Lower Bound

$\mathrm{LB}(\mathrm{H}_N;\theta^u,\beta)=f(\mathrm{h}^*;\theta)=\sum_{i\in N}\hat{\lambda}_i(\mathrm{h}^*)$;

$\hat{\lambda}_i(\mathrm{h})=\theta_i^u(h_i)+\sum_{j\in N_i}\beta_{ji}(h_j,h_i)$

- Define Node-wise-Local-Primal-Dual-Gap (NLPDG)

$\delta_i(\mathrm{h}^*)=\lambda_i(h_i^*)-\hat{\lambda}_i(\mathrm{h}^*)$

- Properties of NLPDG

$\delta_i(\mathrm{h}^*)=\sum_{j\in N_i}\max_{h_j}\beta_{ji}(h_j,h_i^*)-\beta_{ji}(h_j^*,h_i^*)\geq 0$ Non-Negative

$\sum_{i\in N}\delta_i(\mathrm{h}^*)=\mathrm{UB}(\mathrm{H}_N;\theta^u,\beta)-\mathrm{LB}(\mathrm{H}_N;\theta^u,\beta)=\mathrm{PDG}$

Primal-Dual-Gap (PDG) = sum of NLPDG

-Select the Largest NLPDG

PDG = 0 implies the BB search is completed.

Variable State Ordering (VSO)

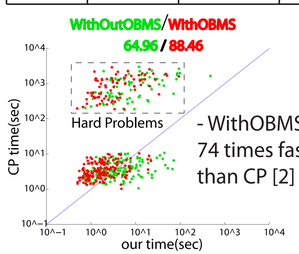
- Split by domain knowledge.
  E.g.: Human Pose:
  location of body part.

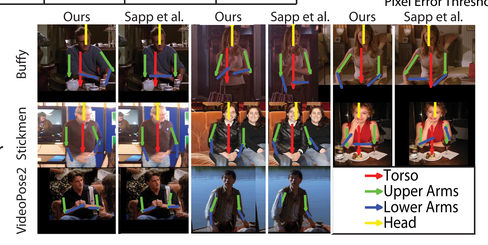
6. Experiment Results

Baseline Methods : CP [2] $O(H^q)$ (q is max cluster size).

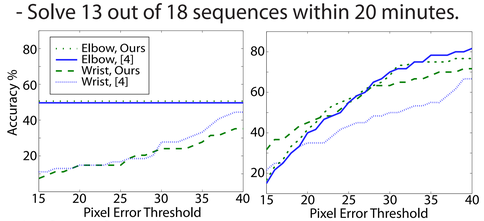
Human Pose Estimation (HPE):
extending a tree-model (CPS) [3] to a fully-connected model.

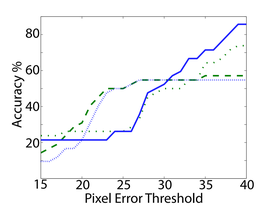
	Buffy				Stickmen	
Parts	Ours F	Ours 13	Ours 7	CPS	Ours F	CPS
Head	**99.15**	**99.15**	**99.15**	**99.15**	**99.44**	99.17
Torso	**99.57**	**99.57**	**99.57**	**99.57**	**99.72**	**99.72**
RUA	**95.30**	93.59	93.16	**95.30**	**82.50**	82.22
LUA	**92.31**	**92.31**	**92.31**	91.88	80.28	**81.67**
RLA	**63.25**	59.83	60.26	59.83	**56.94**	54.44
LLA	**64.53**	62.39	61.97	59.83	**53.89**	51.94





Video Pose Estimation:
Solving Stretchable Model (SM) [4]
- Solve 13 out of 18 sequences within 20 minutes.





7. Conclusion

- Flexible bound for MRFs with different structures (HPE and SM).
- BMT and OBMS are used to speed up from $O(H^2)$ to $O(Q\log_2 H)$.
- Faster than state-of-the-art exact Inference (CP) [2].
- Loopy model achieves superior accuracy than tree model [3].

Acknowledgement

We acknowledge the support of the ONR grant N000141110389, ARO grant W911NF-09-1-0310.

Correspondence

sunmin@umich.edu
muralikt@umich.edu
Honglak@eecs.umich.edu
silvio@eecs.umich.edu

References

[1] A. Globerson and T. Jaakkola.Fixing max-product: Convergent message passing algorithms for MAP LP-relaxations. In NIP'08.
[2] D. Sontag, T. Meltzer, A. Globerson, Y. Weiss, and T. Jaakkola. Tightening LP relaxations for MAP using message-passing. In UAI'08.
[3] B. Sapp, A. Toshev, and B. Taskar. Cascaded models for articulated pose estimation. In ECCV'10.
[4] B. Sapp, D. Weiss, and B. Taskar. Parsing human motion with stretchable models. In CVPR'11.
[5] D. Harel and R. E. Tarjan. Fast algorithms for finding nearest common ancestors. SIAM Journal on Computing, 1984.
[6] M. Sun, M. Telaprolu, H. Lee, and S. Savarese. Efficient and Exact MAP-MRF Inference using Branch and Bound. In AISTATS'12.
[7] M. Sun, M. Telaprolu, H. Lee, and S. Savarese. An Efficient Branch-and-Bound Algorithm for Optimal Human Pose Estimation. In CVPR'12.
[8] P. Felzenszwalb and D. P. Huttenlocher. Pictorial Structures for Object Recognition. In IJCV'05.
[9] D. Tran and D. Forsyth. Improved human parsing with a full relational model. In ECCV'10.
[10] T. P. Tian and S. Sclaroff. Fast globally optimal 2D human detection with loopy graph models. In CVPR'10.
[11] Y. Yang and D. Ramanan. Articulated pose estimation using flexible mixtures of parts. In CVPR'11.
[12] W. Choi, K. Shahid, and S. Savarese. Learning Context for Collective Activity Recognition. In CVPR'11.