# Technical Report: An Efficient Branch-and-Bound Algorithm for Optimal Human Pose Estimation

Min Sun Murali Telaprolu Honglak Lee Silvio Savarese
Department of EECS, University of Michigan, Ann Arbor, MI 48109

## 1. Linear Programming Relaxation

The message passing algorithm [1] defines the edge-wise message as

$$\lambda_{ji}(h_i;\mathcal{H}_j) = \max_{h_j \in \mathcal{H}_j} \beta_{ji}(\hat{h}_j, h_i) \quad (1)$$

which carries information from node $j$ to node $i$. Notice that the edge-wise message $\lambda_{ji}(h_i;\mathcal{H}_j)$ depends on the hypothesis space of node $j$ (i.e., $\mathcal{H}_j$). For conciseness, it is often omitted and the message becomes $\lambda_{ji}(h_i)$. $\lambda_i(h_i)$ is treated as the node-wise message which is the summation of all edge-wise messages into node $i$ (i.e.,$\lambda_{ji}(h_i;\mathcal{H}_j); j \in \mathcal{N}_i$) and the unary potential $\theta^u_i(h_i)$ as defined in Eq. 2 of the main paper. In MP-LP, the edge-wise messages are sequentially updated and a pair of messages $\lambda_{ji}(h_i)$ and $\lambda_{ij}(h_j)$ are updated simultaneously as follows,

$$\lambda_{ji}(h_i) = -\frac{1}{2}\lambda_i^{-j}(h_i) + \frac{1}{2}\max_{h_j \in \mathcal{H}_j}\left(\lambda_j^{-i}(h_j) + \theta_{ij}(h_i, h_j)\right) \quad (2)$$

where $\lambda_i^{-j}(h_i) = \lambda_i(x_i) - \lambda_{ji}(x_i)$ is the sum of messages into node $i$ except the message from node $j$. Notice that when $\lambda_{ji}(h_i)$ is updated, it will eventually update the node-wise message $\lambda_i(h_i)$ by passing the messages into node $i$. As a result, the messages are passed to nodes and further change the other edge-wise messages. The $\beta$s can be retrieved from the messages as follows,

$$\beta_{ji}(h_j, h_i) = -\frac{1}{2}\lambda_i^{-j}(h_i) + \frac{1}{2}\left(\lambda_j^{-i}(h_j) + \theta_{ij}(h_i, h_j)\right) \quad (3)$$

## 2. Learning

The max-margin learning problem is formulated as below,

$$\begin{aligned} \min_{\mathbf{w},\xi\geq 0} \quad & \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_n \xi_n \\ \textit{s.t.} \quad & \forall n, \forall \mathbf{h} \neq \mathbf{h}_n, \quad -\mathbf{w}^T\mathbf{\Psi}(\mathbf{h}, I_n) \geq 1 - \xi_n \\ & \forall n, \quad \mathbf{w}^T\mathbf{\Psi}(\mathbf{h}_n, I_n) \geq 1 - \xi_n \ , \end{aligned} \quad (4)$$

where $h_n$ and $I_n$ are the ground truth part configuration and the image evidence of the $n_{th}$ image, respectively. We use a cutting plane solver [2] to solve the above quadratic programming (QP) problem with a large number of negative constraints (the constraints in the first row). We use the max-margin formulation to learn weights $\mathbf{w}$ such that the ground truth configuration (pose assignment) $h_n$ has the highest score. In other words, the weights are adjusted in such a way that the MAP estimation would become as consistent with the ground truth as possible.

## References

[1] A. Globerson and T. Jaakkola. Fixing max-product: Convergent message passing algorithms for MAP LP-relaxations. In *NIPS*, 2008. 1

[2] I. Tsochantaridis, T. Hofmann, T. Joachims, and Y. Altun. Support vector machine learning for interdependent and structured output spaces. In *ICML*, 2004. 1