IEEE International Conference on Computer Vision 2009 (ICCV'09)

Human Pose Estimation Using Consistent Max-Covering

Hao Jiang
Computer Science Department, Boston College
Chestnut Hill, MA 02467, USA
hjiang@cs.bc.edu

Abstract

We propose a novel consistent max-covering scheme for human pose estimation. Consistent max-covering formulates pose estimation as the covering of body part polygons on an object silhouette so that the body part tiles maximally cover the foreground, match local image features, and satisfy body linkage plan and color constraints. It uses high order constraints to anchor multiple body parts simultaneously; the hyper-edges in the part relation graph are essential for detecting complex poses. Because of using multiple clues in pose estimation, this method is resistant to cluttered foregrounds. We propose an efficient linear relaxation method to solve the consistent max-covering problem. Experiments on a variety of images and videos show that the proposed method is more robust than locally constrained methods for human pose estimation.

1. Introduction

Human pose estimation has been receiving a lot of interest because of its potential important applications in surveillance and human computer interaction. In such applications rough object silhouettes can often be obtained using simple methods such as background subtraction or color segmentation. In this paper, we propose a novel consistent max-covering scheme and efficient linear method to integrate object foreground map with other image features and body plan constraint to achieve reliable pose estimation.

Silhouettes have been used as a direct input in human pose estimation. For simple poses with little self-occlusion, rough body configuration can be extracted by skeleton operation on clean silhouettes [1]. Machine learning methods have also been studied for pose inference using silhouettes [2, 3]. Because of the large variety of human poses, pose regression is a hard problem. Learning methods have to process high dimensional input data and very large training sets. Dimensionality reduction methods such as probabilistic principal component analysis [4], manifold learning [5] and Gaussian process dynamical model [6] have been used to relieve the dimension explosion problem. These methods are currently used for restricted classes of poses and work with clean silhouettes. Shape matching methods have also been used in pose estimation [7, 8, 9]. They match silhouettes or edge maps and are more resistant to clutter. For free-style poses, shape matching methods have high complexity because they have to search through a huge exemplar database.

Apart from the aforementioned top-down approaches, bottom-up pose estimation methods have also been intensively studied. In the bottom-up methods, body part candidates are first detected and then assembled to fit image observations and a body plan. This approach is most related to our method. With silhouettes, an efficient pose estimation method [10] has been studied based on tree model and posterior sampling. This method is resistant to foreground clutter. One difficulty of this formulation is the "over-counting" problem, which happens when multiple body parts occupy the same pixel in an image. Other stochastic searching methods [11, 12] and dynamic programming [13, 14] have also been used in pose optimization. Non-tree methods have recently been proposed to facilitate stronger structure constraints. They are optimized using convex programming [15, 17] or belief propagation [18]. Previous body part assembly methods use local pairwise constraints and do not have mechanisms to control the global structure. For complex poses, local information is not sufficient to yield a correct pose estimation.

In this paper, we follow the body part assembly scheme and propose a novel consistent max-covering method for human pose estimation. In such a formulation, pose estimation becomes the problem of covering an object foreground map with a bunch of body part tiles. These tiles maximally cover the foreground, match image local appearance and are consistent in terms of the body linkage plan and other symmetry constraints. Max-covering with consistent constraint is denoted as *consistent max-covering*. It introduces high order relations among all the body parts since each body part may influence others when forming a covering. The high order body part correlation is essential for finding complex poses when self-occlusions or other part interactions

2009 IEEE 12th International Conference on Computer Vision (ICCV)
978-1-4244-4419-9/09/$25.00 ©2009 IEEE

occur. This method also solves the "over-counting" problem since we explicitly express the overall coverage of a body part assembly. We propose a linear formulation which can be efficiently solved using a relaxation method. Using multiple clues in pose estimation, the proposed method is resistant to occlusion and works with low quality silhouette or soft object foreground mask. Experiments on a variety of images and videos show that the proposed method is robust and efficient in human pose estimation.

2. Pose Estimation by Consistent Max-Covering

Given an object foreground map from color segmentation or background subtraction, pose estimation can be simulated as a jigsaw puzzle problem. In the following, we show how a pose estimation problem can be formulated using the consistent max-covering scheme and how we can solve consistent max-covering using an efficient linear method.

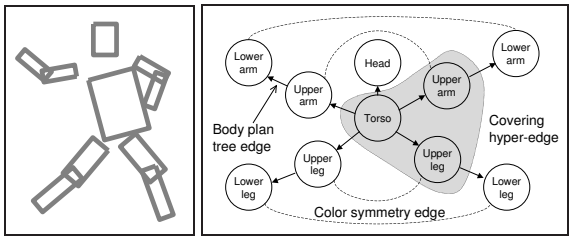


Figure 1. Left: 10-part body model. Right: relation graph of body parts. The gray area shows the example of a hyper-edge.

We use the widely used 10-part body model which contains head, torso, upper arms, lower arms, upper legs and lower legs. Each body part is represented as a rectangle. Our method can also be easily extended to more complex body part shapes. The cardboard figure and the part relation graph are shown in Fig. 1. In our model, the basic body plan follows a tree structure. Apart from interactions between neighboring body parts, consistent max-covering formulation introduces hyper-edges linking body parts that cover the same foreground region, and edges that constrain symmetry body parts. The tree is rooted at the torso and has directional edges. The other two kinds of edges are non-directional.

Similar to other bottom-up pose estimation methods, we first locate potential body part candidates in target images so that we can use them in consistent max-covering. We use simple box detectors to find potential candidates on the edge map of the target image. Chamfer matching is used to match body part templates to the target edge map at different locations and rotations. Non-minimum suppression is then used to locate body part candidates. Since we have a rough foreground map, body part candidates can be further pruned: we only keep the body part candidates whose average foreground potentials are greater than a threshold.

2.1. Consistent Max-Covering

Each body part candidate covers some pixels in the object foreground. Intuitively, the body part tiles should cover foreground pixels as much as possible in a consistent manner. Each pixel in the foreground map corresponds to a floating point number from 0 to 1 which indicates the foreground potential. The higher the potential is, the more likely the pixel belongs to the object foreground. We denote the foreground map as $f_{x,y}$. The consistent max-covering can be formulated as the following optimization problem:

$$\begin{aligned} & \max_{\mathcal{C}}\{\sum_{(x,y)\in I} r_{x,y} - \alpha M(\mathcal{C}) - \beta P(\mathcal{C}) - \gamma S(\mathcal{C})\} \\ s.t. \quad & r_{x,y} = f_{x,y} \text{ if } (x,y) \text{ is covered by parts, else } 0 \\ & \mathcal{C} \text{ is a body part covering} \end{aligned} \tag{1}$$

where $r_{x,y}$ is the covered potential at pixel (x, y) with the current body part covering $\mathcal{C}$: if the pixel (x, y) in the foreground I is covered by body parts, $r_{x,y}$ takes value $f_{x,y}$ and otherwise 0. Therefore, the first term in the objective function equals the overall potential covered by all the body parts. The second term $M(\mathcal{C})$ is the cost of matching the body parts to local image features. The third term $P(\mathcal{C})$ is the degree of the body part configuration following a human body plan. The last term $S(\mathcal{C})$ penalizes the color difference of symmetrical body parts: if the symmetrical parts, e.g. upper arms, have large color difference, S has large value. We reverse the sign of the last three terms so that they are minimized. α, β and γ are positive constants to control the weight among the energy terms. This optimization thus tends to find a consistent max-covering on the object foreground.

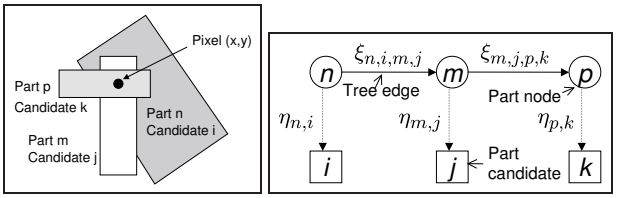
The consistent max-covering in Eqn.(1) is a combinatorial search problem. It is generally NP-hard because of the loopy part relations introduced by covering terms. The large number of feasible covering configurations makes naive exhaustive search infeasible, and for such a problem greedy method is not sufficient. We use a global search method to tackle this problem. In this following, we propose an efficient linear solution.

2.2. Linear Solution

We linearize the consistent max-covering optimization in Eqn.(1) and obtain a mixed integer linear program. It can be further relaxed into a much simpler linear program for efficient solution.

2.2.1 Foreground Covering Potential

In Eqn.(1) the covering term in the objective function is $\sum_{(x,y)\in I} r_{x,y}$, which equals the total potential covered by



(a) (b)

Figure 2. (a) Body part covering example in which part candidates (n,i), (m,j) and (p,k) cover point (x,y) in the foreground map. (b) Definitions of edge variables and node variables.

all the body parts. r represents the covering potential at a pixel. We now explicitly express how the variable r is related to the choice of body part locations.

We introduce a binary indicator variable $\eta_{n,i}$, which is 1 if body part n selects target candidate i and otherwise 0. $\eta_{n,i}$ is therefore a node variable that corresponds to each node of the body graph. Since each body part only has one target location, we need to make sure

$$\sum_{i \in T(n)} \eta_{n,i} = 1, \forall n \in V$$

where $T(n)$ is the target candidate set of body part n; V is the set of all body parts; we also use V to denote the node set of the body graph. We are now ready to specify the covering term using the node variable η. We introduce the following constraint for the covering variable r,

$$\sum_{\forall (n,i) \text{ covers } (x,y)} \eta_{n,i} \geq r_{x,y}$$

in which (n,i) denotes the ith candidate for body part n. We further need to bound $r_{x,y}$ to be a nonnegative number that can be as big as the foreground potential,

$$0 \leq r_{x,y} \leq f_{x,y}$$

For $r_{x,y}$ not covered by any candidate tiles, it is set to zero. As an example shown in Fig. 2(a), there are totally three body part candidates (n,i), (p,k) and (m,j) covering the foreground pixel (x,y). The constraint for $r_{x,y}$ is therefore $r_{x,y} \leq \eta_{n,i} + \eta_{p,k} + \eta_{m,j}$ and $0 \leq r_{x,y} \leq f_{x,y}$.

It is not hard to verify that with the above formulation r equals the covering potential: if at least one body part covers the foreground pixel (x,y), to maximize the objective function in Eqn.(1), $r_{x,y}$ should equal $f_{x,y}$, recalling that body part indicator variable η is 0 or 1 and $f_{x,y}$ is between 0 and 1; if (x,y) is not covered by any part tiles, $r_{x,y}$ will be 0 because of the upper bound constraint. Therefore, given such constraints, $\sum_{(x,y) \in I} r_{x,y}$ is indeed the overall covering potential of the body parts.

2.2.2 Image Matching Cost

Apart from the object foreground covering cost, assigning each body part to a target location involves an image matching cost. In this paper, the cost is the linear combination of the Chamfer matching cost and the local covering cost. Local covering cost equals 1 minus the average covering potential. With the binary node assignment variable η, the total image matching cost can be linearized as:

$$M = \sum_{n \in V, i \in T(n)} c_{n,i} \cdot \eta_{n,i}$$

Here $c_{n,i}$ is the image matching cost of part n at location i.

2.2.3 Spatial Consistency

A good body part assignment should have neighboring body parts linked together: the end points of consecutive body parts are close, and connected limbs have similar orientations. To linearize the spatial consistency function $P(.)$ in Eqn.(1), we introduce an edge indicator variable $\xi_{n,i,m,j}$ for each directional tree edge (n,m); $\xi_{n,i,m,j}$ is 1 if body part n selects candidate i and body part m takes candidate j. The definitions of the edge variable ξ and node variable η are illustrated in Fig. 2(b).

Using edge assignment variable ξ, the total spatial consistency cost is linearized as:

$$P = \sum_{(n,m) \in E, i \in T(n), j \in T(m)} h_{n,i,m,j} \cdot \xi_{n,i,m,j}$$

where E is the tree edge set and the coefficient $h_{n,i,m,j}$ is defined as

$$h_{n,i,m,j} = \begin{cases} ae_{n,i,m,j} & \text{if } n \text{ is torso} \\ ad_{n,i,m,j} & \\ +b \sin^2(\frac{\theta_{n,i} - \theta_{m,j}}{2}) & \text{otherwise} \end{cases}$$

Here $e_{n,i,m,j}$ is the Euclidean distance between a proper end of a torso candidate and the start point of an upper body part candidate; $d_{n,i,m,j}$ is the Euclidean distance between the end point of part n at location i and the start point of part m at location j; $\theta_{n,i}$ is the angle of part n at location i and $\sin^2(\frac{\theta_{n,i} - \theta_{m,j}}{2})$ penalizes large angle difference for consecutive body parts; a and b are positive weight constants.

The node variable η and the edge variable ξ are dependent. The pairwise edge assignment has to be consistent with the node assignment: each body part appearing in different pairs must have a unique assignment of the target candidate. To enforce the assignment consistency, $\forall (n,m) \in E$, we let

$$\eta_{n,i} = \sum_{j \in T(m)} \xi_{n,i,m,j}, \quad \eta_{m,j} = \sum_{i \in T(n)} \xi_{n,i,m,j}$$

Recall that $T(m)$ is the set of covering candidates for part m. The above constraints imply that $\sum_{j \in T(m)} \xi_{n,i,m,j} = \sum_{l \in T(k)} \xi_{n,i,k,l}$, $k \neq m$ and $\sum_{i \in T(n)} \xi_{n,i,m,j} = \sum_{l \in T(k)} \xi_{m,j,k,l}$. This enforces the assignment consistency at common nodes in the tree.

2.2.4 Color Consistency

To linearize the color difference term S in Eqn.(1), we use $L1$ norm to compute the color difference of two body parts so that we can use a standard linear programming auxiliary variable trick. Let H be the set of symmetrical body part pairs. Term S can be linearized as

$$S = \sum_{\{n,m\} \in H} \sum_{k=1}^{3} (g^+_{n,m,k} + g^-_{n,m,k})$$

where $g^+_{m,n,k}$ and $g^-_{m,n,k}$ are nonnegative auxiliary variables. We use 3 color channels $k = 1..3$. The nonnegative auxiliary variables are constrained by the color difference at each channel,

$$g_{n,k} - g_{m,k} = g^+_{n,m,k} - g^-_{n,m,k}, \quad k = 1..3, \quad \forall \{n,m\} \in H$$

Here $g_{n,k}$ is the color of body part n at channel k. The color of a body part can be computed using the node assignment indicator variable $\eta_{n,i}$:

$$g_{n,k} = \sum_{i \in T(n)} l_{n,i,k} \cdot \eta_{n,i}$$

and $l_{n,i,k}$ is the average color of the candidate covering region i for body part n at channel k. It is easy to verify that at least one variable in the pair of $g^+_{n,m,k}$ and $g^-_{n,m,k}$ will become 0, when the objective function is optimized; otherwise we can zero one of them and obtain a better solution by subtracting the two variables with the smaller one of them. Therefore, when the objective function is optimized, $g^+_{n,m,k} + g^-_{n,m,k} = |g_{n,k} - g_{m,k}|$, and S is the $L1$ color distance between symmetrical body parts.

Note that the color symmetry term is a regularization term in the objective function. The constraint of color is therefore soft, which permits occasionally large discrepancy of colors on symmetrical body parts.

2.2.5 Relaxation Solution

The above optimization is a mixed integer program with binary variables ξ, η and continuous variables r and g. Directly solving the mixed integer program has high complexity. We relax it into the following linear program,

$$\begin{aligned}
\max\{ \quad & \sum_{(x,y) \in I} r_{x,y} - \alpha \sum_{n \in V, i \in T(n)} c_{n,i} \cdot \eta_{n,i} - \\
& \beta \sum_{(n,m) \in E, i \in T(n), j \in T(m)} h_{n,i,m,j} \cdot \xi_{n,i,m,j} - \\
& \gamma \sum_{\{n,m\} \in H} \sum_{k=1}^{3} (g^+_{n,m,k} + g^-_{n,m,k}) \} \\
s.t. \quad & \eta_{n,i} = \sum_{j \in T(m)} \xi_{n,i,m,j}, \ \eta_{m,j} = \sum_{i \in T(n)} \xi_{n,i,m,j}, \\
& \xi \geq 0, \quad \forall (n,m) \in E \\
& \sum_{i \in T(n)} \eta_{n,i} = 1, \ \eta_{n,i} \geq 0, i \in T(n), \forall n \in V \\
& g_{n,k} - g_{m,k} = g^+_{n,m,k} - g^-_{n,m,k}, \\
& g^+_{n,m,k}, \ g^-_{n,m,k} \geq 0, \quad \forall \{n,m\} \in H \\
& g_{n,k} = \sum_{i \in T(n)} l_{n,i,k} \cdot \eta_{n,i}, \quad k = 1..3, \quad \forall n \in V \\
& \sum_{\forall (n,i) \text{ covers } (x,y)} \eta_{n,i} \geq r_{x,y}, \\
& 0 \leq r_{x,y} \leq f_{x,y}, \quad \forall (x,y) \in I
\end{aligned}$$

where ξ and η are relaxed into continuous variables in [0,1]. If we do not include the covering constraint term and color symmetry term, the linear program on the tree structure body plan is equivalent to the integer program and it can be solved efficiently using dynamic programming. The non-tree structure of consistent max-covering complicates the solution. Its relaxation does not directly yield integer solutions for node variables η. Direct rounding by selecting the largest η for each body part yields poor results. Fortunately, using the interior method, its solution almost always contains very few large η. We threshold η to zero out most variables. The typical threshold is 0.001. Similar to the approximation trick in [17], we can further construct a small mixed integer program by only including the target candidates corresponding to the nonzero η. The small mixed integer program can be directly solved using simple exhaustive enumeration or the more efficient branch and bound method. Since the first step eliminates a large number of covering candidates, the second exhaustive search step can be quickly solved.

The average complexity of a linear program is roughly linear to the number of constraints and logarithm to the number of variables [19]. This simplex method heuristic applies to our model for which primal-dual interior point method is almost always faster than simplex method for different problem sizes. The number of edge variables of the proposed linear program is proportional to the square of the number of target candidates n; the number of foreground variables equals the number of foreground pixels m. The number of constraints is in the same order as the number of variables. The linear program thus has the $O((n^2+m)\log(n^2+m))$ average complexity. We can further speed up the linear program by heuristics. The neighboring tree nodes only accept quite limited set of candidates: the pair of candidates too close or too far away can be pruned. Using such a trick, the number of edge variables and related constraints can be greatly reduced. The number of foreground variables and constraints can also be reduced by using a coarser representation of the foreground map: instead of corresponding to each pixel, foreground variable and constraint correspond to each region. Due to the slow

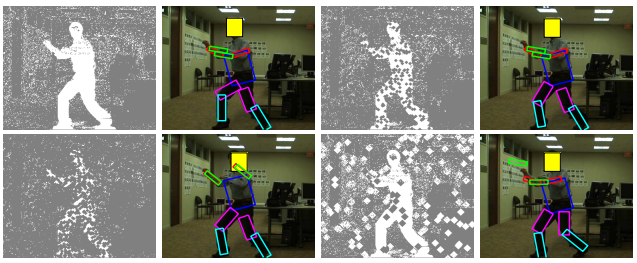


Figure 3. Pose estimation using cluttered foreground. The gray-scale images on the first and third columns are foreground maps with different noisy levels; gray indicates potential 0 and white 1. The pose estimation results are overlapped on the corresponding color images.

Figure 4. Compare with pose estimation using max-covering which ignores the consistency constraints. From left to right: input image, foreground map, pose estimation using max-covering and the result of consistent max-covering.

variation of the foreground map, down sampling will not degrade the performance. Typically we extract 100 torso candidates, 200-500 limb and head candidates, and use 2500-5000 foreground variables to represent foreground map regions. In a 2.8GHz Linux machine, the linear program takes less than 20 seconds to converge.

3. Experimental Results

In this section, we evaluate the consistent max-covering method for human pose estimation and compare the proposed method with different approaches.

Fig. 3 illustrates the proposed method's resistance to cluttered foreground. The foreground map is obtained from background subtraction and shown as the first image in Fig. 3. We add clutter into the foreground map to simulate foregrounds at different noisy levels. As shown in Fig. 3, the consistent max-covering method works well with cluttered foregrounds. The result degrades mildly even though there is substantial amount of clutter in both the object foreground and background.

Fig. 4 illustrates the results of max-covering and consistent max-covering. Max-covering maximizes the covering potential while ignoring other consistency constraints. As expected, max-covering may generate a body part covering that does not resemble a human body plan. Consistent max-covering is necessary to obtain a good result.

We proceed to test another variation of the linear method in which we keep only the local part matching cost and the tree structure spatial consistency constraint. Ignoring the covering energy, this is in fact the formulation which can be solved by dynamic programming (DP) [13]. The proposed relaxation method is exact and equivalent to DP in this case. Without global constraints, we expect more errors. Fig. 5 shows how the proposed method improves the

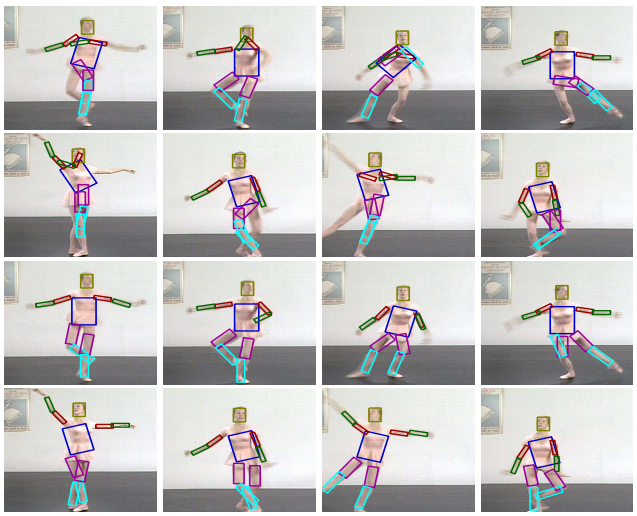


Figure 5. Compare with dynamic programming (DP). The first two rows are the results of DP and the last two rows show how the proposed method improves the result.

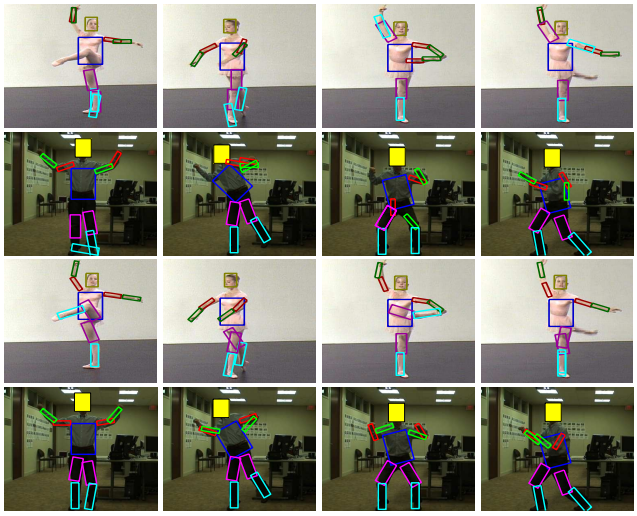


Figure 6. Compare with a simpler linear formulation (Simple LP) which uses only node variables. Rows 1-2 are the results of Simple LP and the last 2 rows show the results of the proposed method.

Table 1. Average Number of Errors Per Frame in Pose Estimation (B1-2, L1-2, T, F indicate ballet, lab, taichi and fitness videos)

	B1	B2	L1	L2	T	F
CM-Covering	0.46	0.74	0.90	0.77	1.29	1.40
DP	3.91	3.47	6.01	3.20	3.30	3.68
Simple LP	0.97	1.43	1.46	1.27	2.31	2.46
Tree [16]	2.80	2.48	1.29	1.09	2.40	1.56
Non-Tree [17]	1.30	1.26	1.51	1.46	2.19	2.08

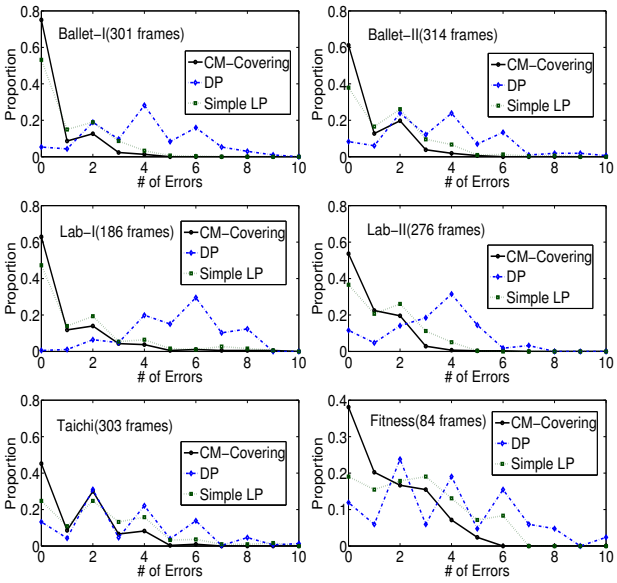


Figure 7. Per-frame error number distributions for different variations of the proposed linear method. CM-Covering is the proposed method; DP ignores covering and color constraints; Simple LP uses only node variables. Good performance is characterized by large portion of a curve at small error range.

result over DP. The complete comparison of the proposed method with DP is based on 6 test video sequences: 2 ballet sequences, 2 lab sequences and 2 videos from YouTube. The ballet sequences include complex movements and body part self-occlusions. In the YouTube sequences and the lab sequences, actors wear baggy clothes and perform complex movements. YouTube videos also have low image quality due to heavy compression. There are totally 1464 images in testing. The comparison using all the 6 test videos is shown in Fig. 7 and Table 1. Fig. 7 shows the normalized histograms of per-frame part detection errors for each video. Since there are 10 body parts, the number of errors per frame is from 0 to 10. The ideal pose detector should have a single peak of 1 at 0 and vanishes anywhere else. A good real detector has an error number histogram focused at the left side and has a short tail at the right side. As shown in Fig. 7, the proposed method is substantially better than DP. The average number of errors per frame in Table 1 confirms the observation, where the errors per frame of the proposed method are less than $1/3$ those of DP.

We further compare the proposed method with a closely related linear programming (LP) formulation. If the body part end point distance is measured with $L1$ norm, we can use the auxiliary variable trick to construct a simpler linear formulation that includes only node variables. In this restrictive case, the proposed method and the simple LP formulation are equivalent with integer constraints. But the relaxation solution has a big difference. As shown in Fig. 6, the proposed method yields better results for challenging cases. The comparison of the two methods using all the 6 test videos is summarized in Fig. 7 and Table 1. The proposed method is consistently better over all the test cases with about half of the per-frame errors.

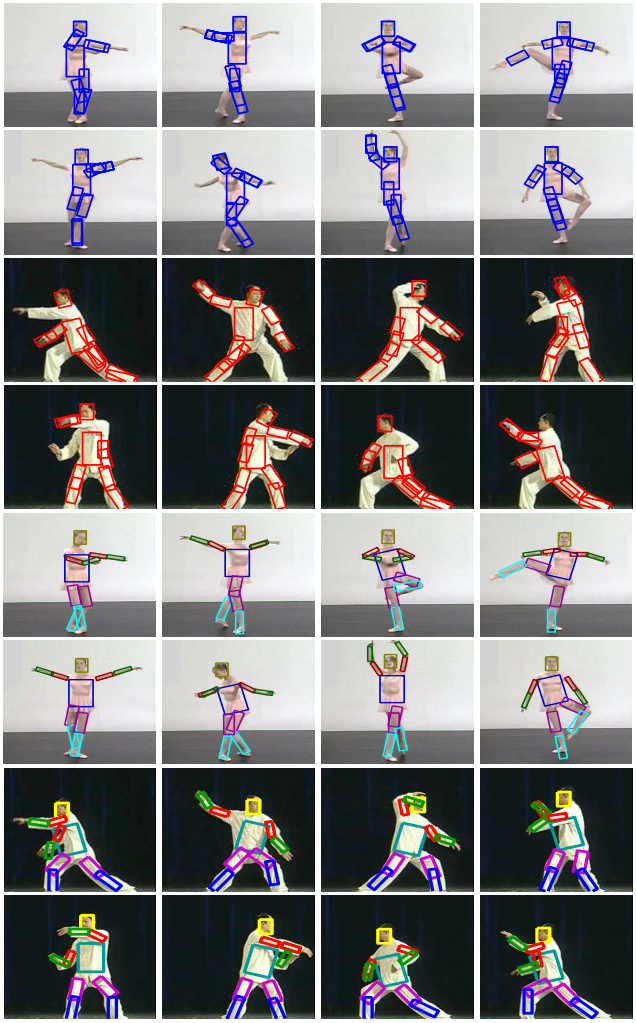


Figure 8. Compare with tree inference method [16]. The first 4 rows are the results using tree inference and the last 4 rows show results of the proposed method.

Our previous tests show that the proposed method is indeed better than its related variations. The question is: does it generate better results than simpler locally constrained methods? One method we compared is the inference method using tree structure [16]. We run the code with this paper on our data. For fair comparison, we use foreground mask to partially eliminate the background clutter before using the code for body part detection. As shown in Fig. 8, tree method sometimes misses body parts because it does not incorporate the global body shape information; the proposed method gives better results. The per-frame error number histogram comparison is shown in Fig. 10 and the comparison of per-frame average error number is in Table 1. The proposed method gives better results over all the test cases. We further compare the proposed method

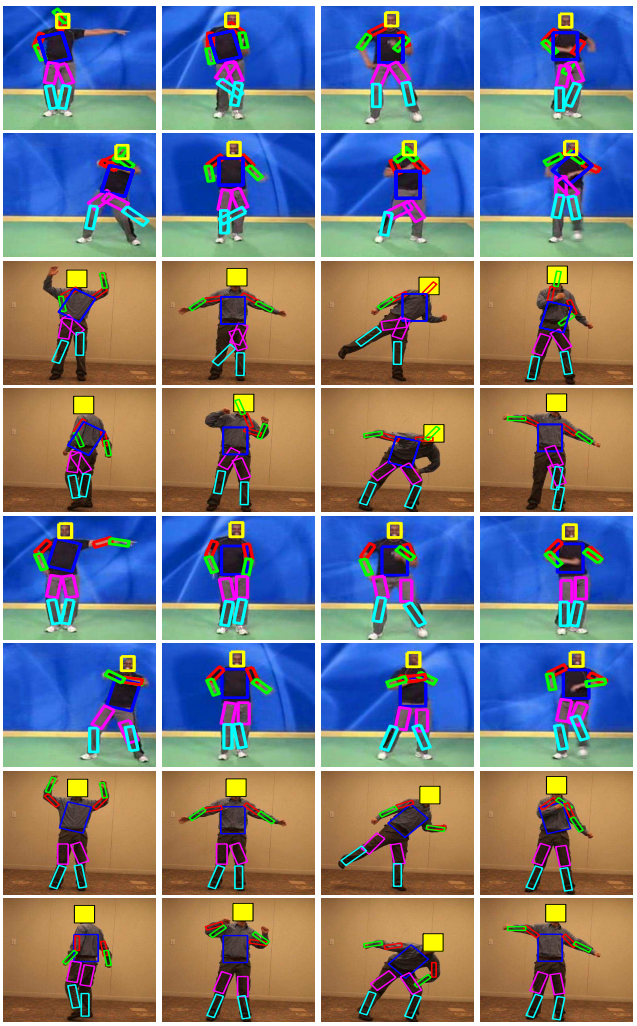


Figure 9. Compare with non-tree method [17]. The first 4 rows are the results using the non-tree method and the last 4 rows show results of the proposed method.

with a non-tree method [17]. This method also uses only local constraints. The results are shown in Fig. 9, Fig. 10 and Table 1. The proposed method still works much better. More sample results of pose estimation using the proposed method are shown in Fig. 11. The proposed method robustly detects body poses in the test sequences. As shown in Fig. 11, a few more errors occur in the taichi and fitness sequences. This is mostly due to the simple body part detector used in this paper. Weak image edges may result in error local matching costs associated with body part candidates. Better body part detector can be used to further improve the results.

4. Conclusion

We propose a novel consistent max-covering scheme for robust human pose estimation. The proposed scheme integrates object overall shape in body part assembly for more

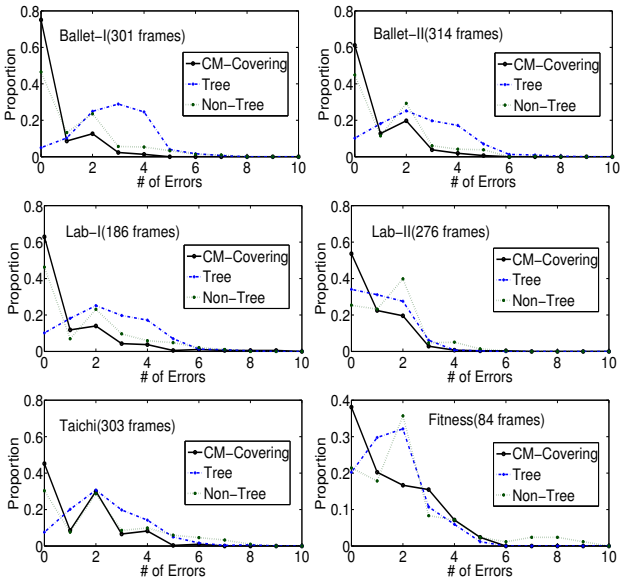


Figure 10. Per-frame error distribution comparison with tree inference (Tree) [16] and non-tree optimization (Non-Tree) [17]. CM-Covering is the proposed method. Good performance is characterized by large portion of a curve at small error range.

robust pose estimation. We combine different clues such as edges, color symmetry, body linkage plan seamlessly and try to find an optimal consistent max-covering of the object foreground map using body part polygons. The proposed method introduces high order correlations among multiple body parts and greatly improves the robustness of pose estimation for complex movements. We propose a linear formulation and an efficient relaxation method. Experiments on challenging video sequences show that the proposed method is robust and efficient in human pose estimation. We believe the proposed method is useful for many applications including automatic surveillance and human movement analysis.

References

[1] X. Bai and W.Y. Liu, "Skeleton pruning by contour partitioning with discrete curve evolution", TPAMI, 29(3), 2007.

[2] R. Rosales and S. Sclaroff, "Inferring body pose without tracking body parts", CVPR 2000.

[3] A. Agarwal and B. Triggs, "3D human pose from silhouettes by relevance vector regression", CVPR 2004.

[4] K. Grauman, G. Shakhnarovich and T. Darrell, "Inferring 3D structure with a statistical image-based shape model", ICCV 2003.

[5] A. Elgammal and C.S. Lee, "Inferring 3D body pose from silhouettes using activity manifold learning", CVPR 2004.

[6] J.M. Wang, D.J. Fleet, A. Hertzmann, "Gaussian process dynamical models for human motion", TPAMI, 30(2), 2008.

[7] G. Mori and J. Malik, "Estimating human body configurations using shape context matching", ECCV 2002.

[8] D.M. Gavrila, "A Bayesian, exemplar-based approach to hierarchical shape matching", TPAMI, 29(8), 2007.

Figure 11. Pose estimation using consistent max-covering for the 6 test videos. `Ballet` has complex movements, self-occlusion, fast motion and interleaving effects; `Lab` involves baggy clothes, cluttered background and foreground; Actors wear baggy clothes in `taichi` and `fitness` sequences from YouTube. We use simple color segmentation to extract foreground maps for all the sequences except `lab-II` for which we use background subtraction. Row 1: sample results for 301-frame `ballet-I`; Rows 2-3: sample results for 314-frame `ballet-II`; Rows 4-5: sample results for 186-frame `lab-I`; Rows 6-7: sample results for 276-frame `lab-II`; Rows 8-10: sample results for 303-frame `taichi`; Row 11-12: sample results for 84-frame `fitness`.

[9] G. Shakhnarovich, P. Viola, and T. Darrell, "Fast pose estimation with parameter sensitive hashing", ICCV 2003.

[10] P.F. Felzenszwalb and D.P. Huttenlocher. "Pictorial structures for object recognition", IJCV, 61(1), 2005.

[11] S. Ioffe and D.A. Forsyth, "Probabilistic methods for finding people", IJCV, 43(1), June 2001.

[12] M.W. Lee and I. Cohen, "Proposal maps driven MCMC for estimating human body pose in static images", CVPR 2004.

[13] P.F. Felzenszwalb and D.P. Huttenlocher, "Efficient matching of pictorial structures", CVPR 2000.

[14] R. Ronfard, C. Schmid, and B. Triggs, "Learning to parse pictures of people", ECCV 2002.

[15] X. Ren, A. Berg and J. Malik, "Recovering human body configurations using pairwise constraints between parts", ICCV 2005.

[16] D. Ramanan, "Learning to parse images of articulated objects", NIPS 2006.

[17] H. Jiang and D.R. Martin, "Global pose estimation using non-tree models", CVPR 2008.

[18] L. Sigal and M.J. Black, "Measure locally, reasoning globally: occlusion-sensitive articulated pose estimation", CVPR 2006.

[19] V. Chvátal, Linear Programming, W.H. Freeman and Company, New York, 1983.