Learning and Matching Line Aspects for Articulated Objects

Xiaofeng Ren
Toyota Technological Institute at Chicago
1427 E. 60th Street, Chicago, IL 60637
xren@tti-c.org

Abstract

Traditional aspect graphs are topology-based and are impractical for articulated objects. In this work we learn a small number of aspects, or prototypical views, from video data. Groundtruth segmentations in video sequences are utilized for both training and testing aspect models that operate on static images.

We represent aspects of an articulated object as collections of line segments. In learning aspects, where object centers are known, a linear matching based on line location and orientation is used to measure similarity between views. We use K-medoid to find cluster centers. When using line aspects in recognition, matching is based on pairwise cues of relative location, relative orientation as well adjacency and parallelism. Matching with pairwise cues leads to a quadratic optimization that we solve with a spectral approximation. We show that our line aspect matching is capable of locating people in a variety of poses. Line aspect matching performs significantly better than an alternative approach using Hausdorff distance, showing merits of the line representation.

1. Introduction

How to represent shape of an object so as to facilitate reliable recognition is a perpetual question in vision. Psychophysical evidences [24] show that, rather than 3D structures as introspection would suggest, human vision system uses view-specific representations. Such discoveries have motivated the use of 2D view-based models in computer vision, a classical example being the *aspect graph* of Koenderink and van Doorn [18].

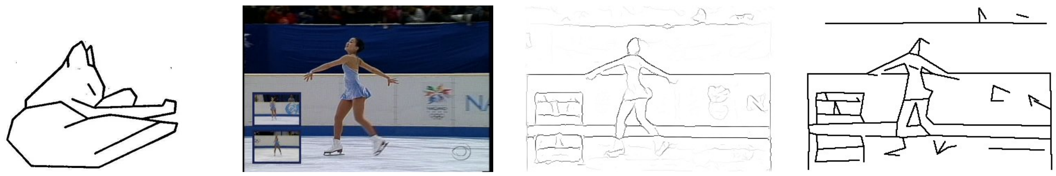
The concept of aspect graphs builds on the insight that the infinitely many possible 2D views of a 3D object may be organized into a small number of groups, or *aspects*. Views in the same aspect are similar to each other and may be easily matched. The similarity "measure" in the classical aspect graph is the topology of the 2D view. Computing and enumerating aspects exactly is usually very hard. For articulated objects, as the figure skating examples in Figure 1 illustrate, variations in pose would generate an aspect graph too large to be practical.

Figure 1. 2D Shape of an articulated object varies greatly because of pose change. We utilize video sequences of people with groundtruth segmentation to learn a small set of view-based models, or *aspects*, and apply them to detecting people in static images.

Koenderink's insight, nevertheless, still holds for articulated objects. The huge number of possible views are not independent; they are closely related to one another and form groups or clusters. Comparing to a nearest neighbor matching strategy that keeps a large number of exemplars in memory, a small number of aspects or view-based templates would show a better understanding of the problem, and be more efficient in practice as well. In this work, we utilize video with groundtruth segmentation to learn a small number of aspects for articulated objects, and to learn how to match them to images.

2D template matching has been in fact a popular paradigm in computer vision. For articulated object detection and localization, Gavrila and Philomin [15] used a large number of stored exemplars and the Chamfer distance metric. Huttenlocher et al [17] used the more robust Hausdorff metric for matching, and Felzenszwalb [10] showed that learning templates under the Hausdorff distance may be viewed as learning half-space models. Toyama and Blake [28] learned templates for people in a tracking setting. Using the Chamfer or Hausdorff metric is common in these approaches, and generally a large number of exemplars are required to cover pose variations.

The novelty of our approach lies in the representation of aspects or 2D templates as collections of line segments. A line representation has many advantages over a point rep-



(a) (b) (c) (d)

Figure 2. A line representation of images. (a) Attneave's Cat (1954), illustrating the observation that shape may be well approximated by piecewise straight lines [1]. Following this observation, we construct a line representation of images from bottom-up: (b) an image, (c) a boundary map computed with the *Probability-of-Boundary* operator [22]. We construct a line approximation (d) by tracing contours in (c) and recursively splitting them until each piece is approximately straight [26].

resentation: it is compact and perceptually meaningful; it allows one to define and use mid-level cues such as adjacency or parallelism; and it naturally corresponds to parts in articulated objects, which commonly appear as pairs of parallel line segments.

To explore the merits and potentials of a line representation, in this work we restrict ourselves to line-based geometric cues. We show how a line representation enables us to learn a small number of aspects for articulated objects, and how line aspect matching may be solved efficiently, leading to promising experimental results.

2. Line Aspects

The use of line representations may be traced to the theory of Attneave [1] . Figure 2 (a) shows Attneave's famous sleeping cat figure. Attneave made an important observation, in 1954, that the shape of an object may be well approximated by line segments, if one places landmarks at extrema of curvature and connects the landmarks by straight lines. His cat does look vividly like a cat and we have no trouble recognizing it. This theory of line approximation has been supported by psychophysical experiments.

The advantages of a line representation have motivated many line-based approaches in computer vision. At a time when line drawings were the primary subjects of study, line representations were popular and successful, especially in recognizing and aligning 3D polyhedral objects (e.g. [20]). The work of Bergevin and Levine [5] showed an example of how people build generic shape models based on line representations. Line representations have been used in other problems such as recovering scene geometry of man-made structures (e.g. [9]) or the camera geometry between views (e.g. [2, 27]). Smooth curves, instead of straight lines, were also studied as an underlying representation of shape (e.g. [21]).

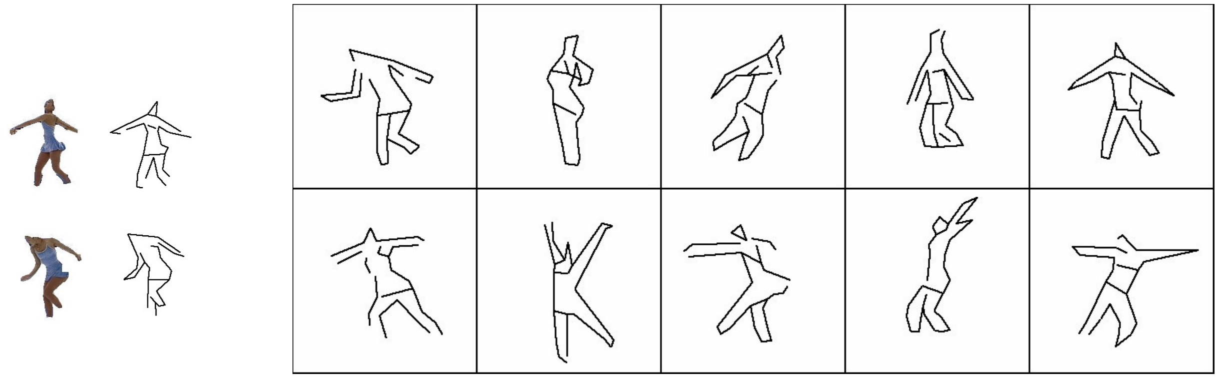
Matching line representations for object recognition is inherently challenging. A line segment itself is non-distinctive; if we allow rotational invariance a line may be matched to any other line. The distinctiveness of a line representation lies in the inter-relationships between the segments. This easily leads to a graph matching problem, known to be hard in general. Line matching was typically solved in a greedy fashion [14] but it did not work well in cluttered scenes. Huet and Hancock [16] approximated all-pair relations with local N-nearest neighbor descriptions. Beveridge and Riseman [6] explored the use of local search. Fleuret and Geman [13] used a coarse-to-fine matching strategy. The recent contour segment network model of Ferrari et al [12] matched rigid objects by searching over paths through contours.

The computational challenges of dealing with pairwise relations have pushed most modern shape matching approaches away from a line representation. Many have adopted a point set representation and could get around the quadratic matching problem by building rich local descriptors, such as the Shape context descriptor [3], to make point features distinctive by themselves. Due to the lack of geometric structure in these point set representations, however, for articulated objects often a large number of templates are needed. Part-based models (e.g. [11]) are another popular approach for articulated objects; they are not view-based and generally require high-level knowledge of the parts and a model relating them.

Meanwhile, progress has been made in graph matching and integer quadratic problems (IQP). Berg et al [4] employed a linear approximation to the IQP problem and showed that using pairwise constraints improved general shape matching. Ren et al [25] adopted the linear approximation strategy in solving human body pose estimation using all-pair relationships. Leordeanu and Hebert [19] and Cour et al [7] developed efficient spectral approximations to the graph matching problem.

Following Attneave's observation, we build a line representation of an image from bottom-up. We apply the *Probability-of-Boundary* operator [22] to compute a soft boundary map that summarizes brightness, color and texture contrasts. Then we trace the contours in the image and recursively partition the contours at high-curvature locations until each piece is approximately straight [26]. Figure 2 (b)-(d) shows an example of this construction.

We use video sequences with groundtruth segmentation for learning aspects. With object centers aligned in the



(a) (b)

Figure 3. Learning line aspects using groundtruth segmentation. (a) Examples of the training data, views of a figure skater, masked by a groundtruth segmentation. A linear matching measures the similarity between views, each represented as a collection of line segments. (b) A K-medoid algorithm is applied to obtain a set of 10 cluster centers, or *aspects*.

training data, we use a linear program based on location and orientation to solve assignments or correspondences between sets of line segments. Correspondences provide similarity scores, and the K-medoid algorithm is used to find a small number of prototypical poses, or *aspects*.

Matching line aspects to novel images is more challenging as object center is unknown, and the image may contain a fair amount of clutter. We rely on relative or pairwise geometric cues, including relative location, relative orientation, and consistencies in adjacency and parallelism. Relative cues lead to a quadratic program, which we solve using a spectral approximation followed by a greedy discretization. We make use of temporal ordering in video to learning feature weights in an iterative procedure.

We again use video with groundtruth segmentation to evaluate the performance of line aspect matching. We compare to two other approaches, an alternative aspect learning approach using the Hausdorff metric [17, 10], and the *Histogram-of-Oriented-Gradients* (HOG) pedestrian detector of Dalal and Triggs [8]. We show that line aspect matching performs the best, successfully finding and localizing people under pose variation and clutter, despite being restricted to simple geometric cues.

3. Learning Line Aspects

Assume we have a collection of images for an articulated object together with groundtruth segmentation. Each image $I^{(k)}$ is represented as a set of line segments. Using the segmentation mask, we extract a view $V^{(k)}$, as a set of line segments $\{l_i^{(k)}\}$, clean from background clutter. The segmentation also provides the center of the object. Hence we can align the views $\{V^{(k)}\}$ such that their centers coincide.

To learn a small set of aspects from these views, we need to first define a similarity metric. We compute a correspondence or assignment between pairs of views. Because the views are aligned, this is not a hard problem and straightforward to do: corresponding lines should be close both in spatial location and orientation.

One issue with a line-based representation is that bottom-up line construction is never perfect. A smooth contour could be a single segment in one view but split into two or more segments in another. Therefore a one-to-one correspondence is not sufficient. We need to accommodate "fractional" assignments between lines.

Let $\{l_i^{(1)}\}$ and $\{l_j^{(2)}\}$ be two sets of line segments to be matched, and let x_{ij} be the mass or number of pixels assigned from each line $l_i^{(1)}$ to $l_j^{(2)}$. Let $b_i^{(1)}$ be the length or mass of lines in view 1, and $b_j^{(2)}$ the length of lines in view 2; they are normalized such as $\sum_i b_i^{(1)} = \sum_j b_j^{(2)}$. We use two features: d_{ij}, the minimum distance between two line segments $l_i^{(1)}$ and $l_j^{(2)}$, and θ_{ij}, the orientation difference between them. We solve a linear assignment problem:

$$\min L(x) = \sum_{i,j} x_{ij}\left(w_d d_{ij} + w_\theta \theta_{ij}\right) \tag{1}$$

$$\text{s. t.} \sum_j x_{ij} = b_i^{(1)}, \sum_i x_{ij} = b_j^{(2)}, \quad x_{ij} \geq 0$$

The minimum cost L^* is the distance (dissimilarity) between the two views [1].

Once we have computed distances L^* between all pairs of views, we run the K-medoid algorithm to find cluster centers. Figure 3 shows the 10 aspects we obtain through a clustering of 375 views from a figure skating video. We will use these 10 aspects in our matching experiments.

[1] w_d and w_θ are set by hand; in our experiments we find that the matching and clustering processes are fairly robust w.r.t. these weights.

4. Matching Line Aspects

In the last section we have found a small set of aspects $T^k = \{l_i^k\}$, or prototypical templates, through linear assignments and K-medoid clustering. We now discuss how we apply these aspects and match them to novel images.

4.1. Pairwise Geometric Cues

It is much more challenging to match an aspect $T = \{l_i\}$ to an image $I = \{l'_j\}$, both represented as collections of line. We do not know the object center, and the image could contain a fair amount of background clutter. Consider a single assignment $l_i \rightarrow l'_j$: without knowing the object center, it becomes meaningless to measure the *absolute* distance between $l_i \in T$ and $l'_j \in I$.

On the other hand, it is useful to measure the *relative* distance between two pairs of assignments. If two lines in the aspect are close to each other, their corresponding lines in the image should also be close. Relative orientation conveys information in a similar way.

For a pair of assignments $l_1 \rightarrow l'_1$ and $l_2 \rightarrow l'_2$, we use a set of simple geometric cues:

1. $f_{d,m}$, *relative location (minimum)*: consider the pair l_1 and l_2 in T, and let $\vec{p}_{11} \in l_1$ and $\vec{p}_{12} \in l_2$ be the pair of points that minimizes the distance between l_1 and l_2 in T. Similarly, let $\vec{p}_{21}$ and $\vec{p}_{22}$ be the two points that minimize the distance between l'_1 and l'_2 in I. We define $f_{d,m} = |(\vec{p}_{11} - \vec{p}_{12}) - (\vec{p}_{21} - \vec{p}_{22})|$;

2. $f_{d,c}$, *relative location (center)*: let $\vec{c}_{11}$, $\vec{c}_{12}$ be the centers of l_1 and l_2, and $\vec{c}_{21}$, $\vec{c}_{22}$ the centers of l'_1 and l'_2. We define $f_{d,c} = |(\vec{c}_{11} - \vec{c}_{12}) - (\vec{c}_{21} - \vec{c}_{22})|$;

3. $f_{d,\theta}$, *relative location (angle)*: $f_{d,\theta}$ is defined as the angle between the two displacement vectors $(\vec{c}_{11} - \vec{c}_{12})$ and $(\vec{c}_{21} - \vec{c}_{22})$;

4. f_θ, *relative orientation*: all four line segments have an orientation. Let $\theta_1 \in \{0, \pi/2\}$ be the difference in orientation between l_1 and l_2, and θ_2 the difference between l'_1 and l'_2, we let $f_\theta = |\theta_1 - \theta_2|$;

5. f_A, *adjacency*: adjacency is a case of relative location that merits special attention. Let $d_1 = |\vec{p}_{11} - \vec{p}_{12}|$ and $d_2 = |\vec{p}_{21} - \vec{p}_{22}|$, we use two thresholds $\tau_l < \tau_h$ to detect violations when one pair of lines are adjacent and the other pair are not: $f_A = (d_1 < \tau_l)(d_2 > \tau_h) + (d_1 > \tau_h)(d_2 < \tau_l)$;

6. f_P, *parallelism*: similarly, as a special case of relative orientation, we define a parallelism cue $f_P = (\theta_1 < \xi_l)(\theta_2 > \xi_h) + (\theta_1 > \xi_h)(\theta_2 < \xi_l)$ with a pair of thresholds $\xi_l < \xi_h$.

Note that we restrict ourselves to pairwise cues only in the matching model. These cues are purely based on the geometry of two pairs of corresponding line segments; each of them incurs a cost if it detects a discrepancy in the pair of correspondences. We combine these cues linearly into a single pairwise cost C:

$$C(l_{i1}, l'_{j1}; l_{i2}, l'_{j2}) = \sum \alpha_k f_k(l_{i1}, l'_{j1}; l_{i2}, l'_{j2}) \quad (2)$$

This pairwise cost leads to a quadratic matching problem.

4.2. Quadratic Matching

To find the best match under the pairwise cues introduced above is a quadratic optimization problem. Suppose we want to find a match between an aspect $T = \{l_1, \cdots, l_n\}$ and an image $I = \{l'_1, \cdots, l'_m\}$. Due to the asymmetry of the matching, i.e. every line in T should have a match if possible but most lines in I will not, we set up the assignment variables $\{x_{ij}\}$ slightly differently from that in Section 3: we now define x_{ij} as the *fraction* of the line $l_i \in T$ that is assigned to $l'_j \in I$ in the matching. We also introduce an outlier dummy node l_0, which may be used to collect clutter lines in I. Any pair of assignments involving l_0 has a constant cost c_0.

Let $\mathbf{x} = \{x_{ij}\}$ be the assignment vector of length $(n+1)m$, and let $\mathbf{W}$ a matrix of size $(n+1)m$-by-$(n+1)m$ whose entries are

$$W(i_1, j_1, i_2, j_2) = \max\left(-C(i_1, j_1, i_2, j_2) + c_0, 0\right)$$

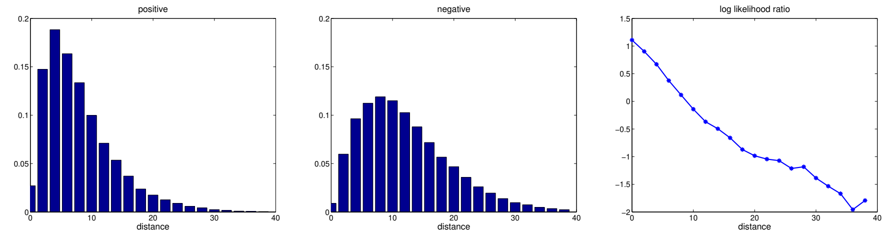
where c_0 is the outlier cost. Let $\{b_i\}$ and $\{b'_j\}$ be the length of line segments in T and I respectively. The best match $\mathbf{x}^*$ solves the following quadratic problem:

$$\begin{aligned} &\max \;\; Q(\mathbf{x}) = \mathbf{x}^T \mathbf{W} \mathbf{x} \\ &\text{s.t.} \sum_{j=1}^{m} x_{ij} = 1; \; \sum_{i=1}^{n+1} b_i x_{ij} = b'_j; \; x_{ij} \geq 0 \end{aligned} \quad (3)$$

This is not an integer program, as we allow fractional assignments. However it is still hard to solve exactly, given all the constraints. A typical problem size has $m \approx 30$ and $n \approx 150$ in our experiments. We use the spectral approximation techniques recently developed for graph matching and quadratic assignment problems [19, 7]. In the spectral approximation, constraints are ignored, and we find the leading eigenvector $\mathbf{x}^*$ of $\mathbf{W}$, corresponding to the largest eigenvalue λ^* [2]. We select the aspect that maximizes $Q(\mathbf{x}^*)$ normalized by n.

We construct a fractional matching $\{x_{ij}\}$ from $\mathbf{x}^*$ by a greedy procedure: first we set a constant threshold on $\mathbf{x}^*$. Then we sort all the plausible matchings above threshold,

[2] As pointed out in [19], the Perron-Frobenius theorem guarantees that $\mathbf{x}^*$ has non-negative entries.



(a) (b) (c)

Figure 4. We train model parameters iteratively. Shown here in (a) and (b) are empirical distributions of one feature, $f_{d,c}$, the difference in relative center location for a pair of matched line segments. Empirical distributions for both positive and negative cases follow a gamma distribution; the log likelihood ratio is approximately linear (c). This motivates combining features linearly.

and go through them in the order of decreasing x^* value. We maintain the amount of mass not yet assigned in both sets of lines; when considering a plausible matching (i, j), we assign from l_i to l'_j the maximum possible mass (i.e. minimum of the mass remaining unassigned in the two lines).

From the matching $\{x_{ij}\}$, we can estimate the object center. Assume the line aspect T is centered at the origin. Let c_i be the center of line l_i in T, and c'_j the center of line l'_j in the image I. Object center in the image can be estimated as a weighted average:

$$O^* = \left(\sum_{i,j} x_{ij}(c'_j - c_i) \right) / \sum_{ij} x_{ij}$$

4.3. Learning Matching Parameters

We learning the parameters $\{\alpha_k\}$ in the linear cue combination by making use of the fact that we have a video sequence for training, not just a random collection of images. Temporal coherence in video implies that objects are likely to have similar poses in frames that are close in time.

This observation leads to an iterative discriminative learning strategy. Each frame $I^{(k)}$ is represented as a set of lines, and the groundtruth segmentation gives us a (segmented) view $V^{(k)}$. Given a tentative set of weights $\{\alpha_k\}$, we match $V^{(k)}$ to images $I^{(k+d)}$ for $|d| < d_0$, and use these matchings as positive examples. We collect feature values f for all pairs of lines that are matched in these examples. Similarly we match $V^{(k)}$ to images $I^{(k+d)}$ with $|d| > d_1 \gg d_0$, and collect feature values from these negative examples.

We treat the features $\{f_k\}$ as independent, and learn each weight α_k independently from the positive and negative examples. Each α_k is estimated with a 1-D logistic regression. Figure 4 shows an example of the empirical distributions. We find that this procedure converges quickly, after a few iterations, under our experimental setting.

5. Experiments

We utilize two figure skating sequences for our experiments, one Michelle Kwan sequence of 750 frames, and one Tara Lipinski sequence of 1410 frames. The images are of resolution 240-by-360. Groundtruth segmentations are available for both sequences, obtained through tracking. We use half of the Michelle Kwan sequence (375 frames) for learning the aspects and training the parameters. 10 aspects are obtained, as shown in Figure 3.

The pairwise line matching approach is then tested on the second half of the Kwan sequence as well as the Lipinski sequence, independently on each frame. We match every aspect to the images and pick the best using the match score $Q(\mathbf{x})$. We estimate object centers from groundtruth segmentations, and use these centers for evaluation.

We compare to two alternative approaches: one comparison is that we keep the same setup for learning and matching aspects, but replacing line-based matching with the Hausdorff distance metric [17, 10] on 2D pixel templates [3], with or without using edge orientation. We use a robust rank value of 80% in Hausdorff matching. Again, 10 aspects are learned and used for testing.

The second comparison is to a state-of-the-art pedestrian detector of Dalal and Triggs [8] , the *Histogram-of-Oriented-Gradients* (HOG) detector. We used their pretrained parameters, tuned for pedestrian detection, and lowered thresholds to force the detector fire on most images .

We evaluate performance by measuring the distance between estimated object centers and groundtruth. In the last section we have discussed how to estimate object center from a line matching. The Hausdorff detector as well as the Dalal-Triggs pedestrian detector return an object center directly, center of a bounding box. We count the percentage of matches that fall within a distance threshold of the groundtruth center, with the threshold varying from 5 to 40 pixels. The results are shown in Figure 5.

As we can see from these results, detecting and local-

[3]Preliminary experiments show that the Hausdorff metric performs slightly better than the Chamfer metric.

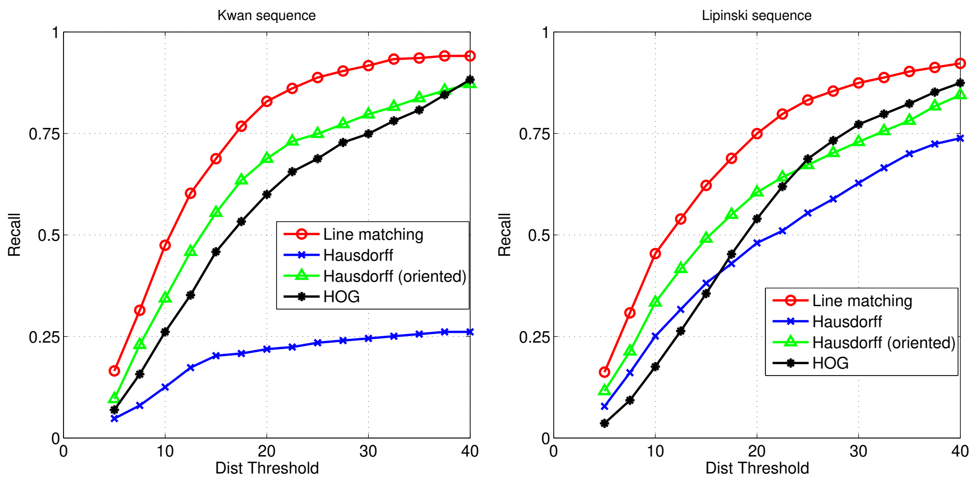
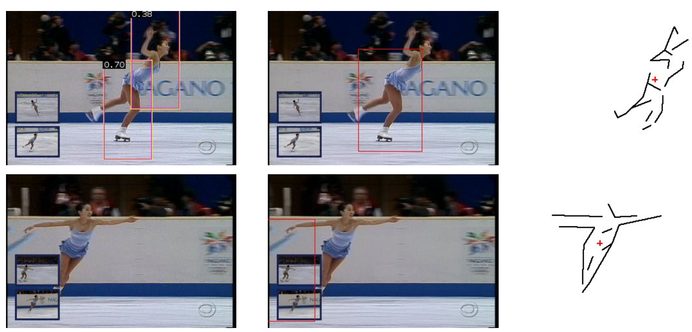


Figure 5. Quantitative comparison of four approaches tested: our line aspect matching, the HOG pedestrian detector [8], and Hausdorff aspect matching [17] with and without orientation. Segmentation provides groundtruth object centers. We evaluate performance by measuring *recall*, or the percentage of correct localizations, under a varying distance threshold. Line aspect matching performs the best on both skating sequences, averaging about 10% to 15% improvement over Hausdorff matching, showing the merits of a line representation.



(a) (b) (c)

Figure 6. Examples comparing line aspect matching (column c) to the HOG detector (column a) as well as Hausdorff matching with orientation (column b). The HOG detector, tuned for pedestrians, have difficulties handling pose variation in this setting. It locates the skater in a wrong place in the first example, and does not fire in the second example. Hausdorff matching does not localize the skater well in the first example, and gets distracted to clutter in the second example. In comparison our line aspect matching locates the skater precisely and recovers most of the object boundaries.

izing the skater in these sequences is not easy, mostly due to large variations in pose. Using edge maps without orientation information, naïve Hausdorff template matching fails badly. Hausdorff matching performs much better when using orientation information and grouping edge elements into orientation channels. The HOG detector, tuned to detecting pedestrians in common poses, does not perform very well on this task. Line aspect matching performs the best on both test sequences, averaging about 10% to 15% higher in accuracy over the second best, Hausdorff matching with orientation. Examples for comparison are shown in Figure 6.

As other settings are identical comparing to the Hausdorff aspects, these results clearly suggest that a line-based representation is preferable for articulated objects. A not-so-surprising observation perhaps, the nature of articulated objects is such that they consist of moving parts, and parts are best characterized by straight lines that connect at joints.

We show more line aspect matching results in Figure 7. Our line aspect matching reliably locates the skaters under large pose variation and background clutter, and recovers most of the object boundaries. It performs surprisingly well, considering the limited amount of information made available to it. Matching is purely based on geometric cues , no color or gradient information is available.

The matching is not perfect; some boundaries are missing, and the pose of the best matched aspect may not correspond well to that of the skater . Experiments show that if we do a second-round matching, making use of the rough knowledge of object center, more complete boundaries may be recovered. Localization performance increases only marginally, however; hence we choose to keep the story simple and use pairwise (relative) cues only.

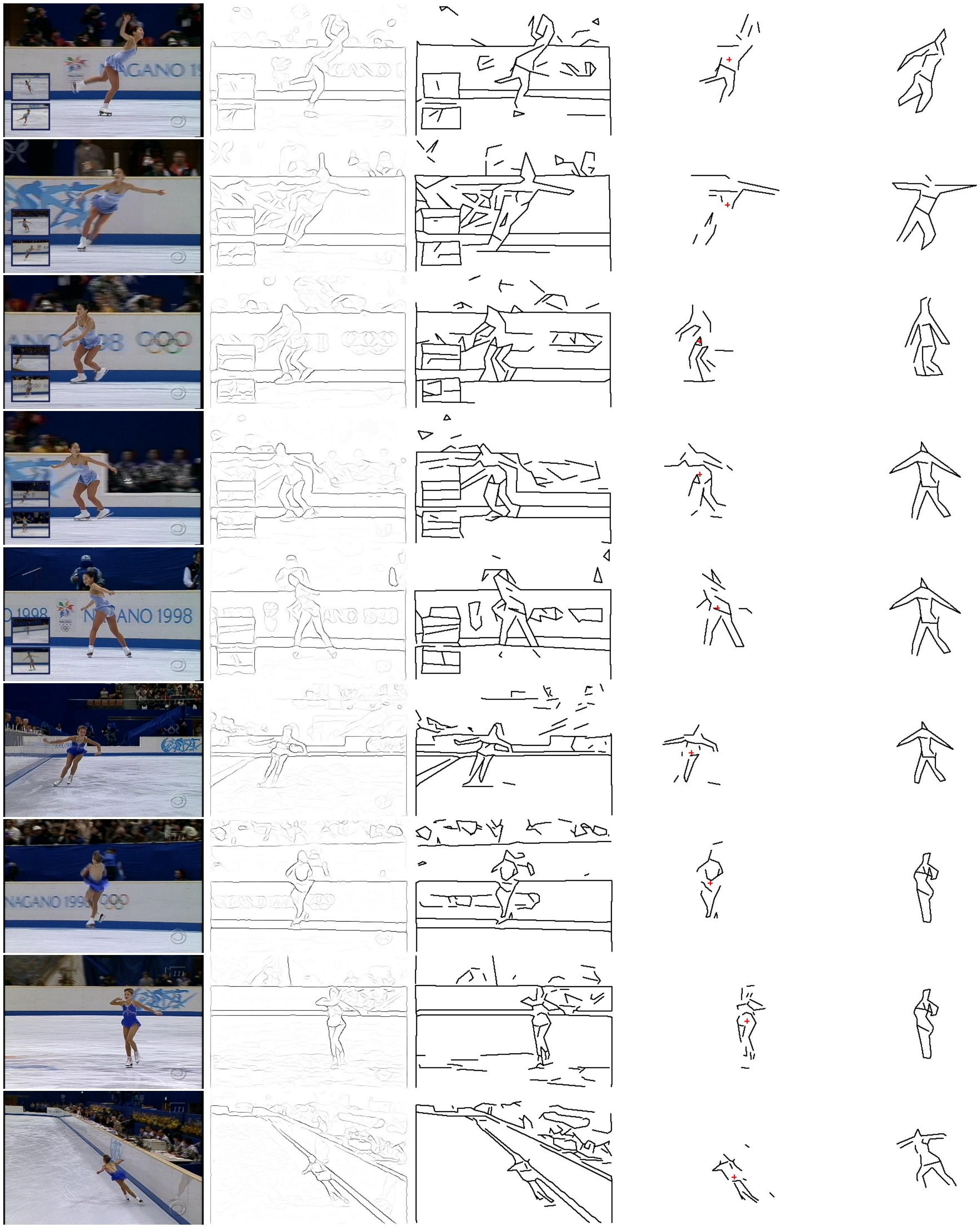


Figure 7. Pairwise line aspect matching results on two skating sequences. Matching is purely edge-based and done independently in each frame, no motion or temporal coherence used. In 5 columns we show the original image, the boundary map computed, the line approximation of boundaries (input to our matching algorithm), matched lines with the estimated center, and the best matched aspect.

6. Discussion

In this work we have utilized video data with groundtruth segmentation to learn a small number of view-based models, or *aspects*, for detecting people. We represent aspects as well as images as a collection of line segments, and use a simple set of pairwise geometric cues for matching. Pairwise cues lead to a quadratic problem that is solved with a spectral approximation. We show that line aspect matching performs well on skating sequences with large pose variations. It performs much better than an alternative approach using Hausdorff metric, suggesting a line-based representation is preferable for articulated objects.

Relative geometric cues only are used in this work. This is intentional as we want to understand the merits and potentials of the line representation. The results are fairly encouraging, and it would be interesting to explore the combination of line geometry with other sources of information, such as color, texture or motion, or to include rich unary shape descriptors into the pairwise matching framework.

References

[1] F. Attneave. Some informational aspects of visual perception. *Psychological Review*, 61:183–193, 1954. 2

[2] N. Ayache and B. Faverjon. Efficient registration of stereo images by matching graph descriptions of edge segments. *Int'l. J. Comp. Vision*, 1(2), 1987. 2

[3] S. Belongie, J. Malik, and J. Puzicha. Matching shapes. In *ICCV*, volume 1, pages 454–461, 2001. 2

[4] A. Berg, T. Berg, and J. Malik. Shape matching and object recognition using low distortion correspondence. In *CVPR*, volume 1, pages 26–33, 2005. 2

[5] R. Bergevin and M. D. Levine. Generic object recognition: building and matching coarse descriptions. *IEEE Trans. PAMI*, 15(1):19–36, 1993. 2

[6] J. R. Beveridge and E. M. Riseman. How easy is matching 2d line models using local search. *IEEE Trans. PAMI*, 19(6), 1997. 2

[7] T. Cour, P. Srinivasan, and J. Shi. Balanced graph matching. In *NIPS*, 2006. 2, 4

[8] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, volume 1, pages 886–893, 2005. 3, 5, 6

[9] P. Debevec, C. J. Taylor, and J. Malik. Modeling and rendering architecture from photographs: A hybrid geometryand image-based approach. In *SIGGRAPH*, 1996. 2

[10] P. Felzenszwalb. Learning models for object recognition. In *CVPR*, 2001. 1, 3, 5

[11] P. Felzenszwalb and D. Huttenlocher. Pictorial structures for object recognition. *Int'l. J. Comp. Vision*, 61(1):55–79, 2005. 2

[12] V. Ferrari, T. Tuytelaars, and L. V. Gool. Object detection by contour segment networks. In *ECCV*, 2006. 2

[13] F. Fleuret and D. Geman. Coarse-to-fine face detection. *Int'l. J. Comp. Vision*, 41(1-2):85–107, 2001. 2

[14] H. Freeman. Computer processing of line-drawing images. *ACM Computing Surveys*, 6(1):57–97, 1974. 2

[15] D. Gavrila and V. Philomin. Real-time object detection for smart vehicles. In *ICCV*, pages 87–93, 1999. 1

[16] B. Huet and E. R. Hancock. Line pattern retrieval using relational histograms. *IEEE Trans. PAMI*, 21(12):1363–1370, 1999. 2

[17] D. Huttenlocher, G. Klanderman, and W. Rucklidge. Comparing images using the Hausdorff distance. *IEEE Trans. PAMI*, 15(9):850–863, 1993. 1, 3, 5, 6

[18] J. J. Koenderink and A. J. van Doorn. The internal representation of solid shape with respect to vision. *Biological Cybernetics*, 32:211–216, 1979. 1

[19] M. Leordeanu and M. Hebert. A spectral technique for correspondence problems using pairwise constraints. In *ICCV*, volume 2, pages 1482–1489, 2005. 2, 4

[20] D. Lowe. Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence*, 31(3):355–395, 1987. 2

[21] J. Malik. Interpreting line drawings of curved objects. *Int'l. J. Comp. Vision*, 1(1):73–103, 1987. 2

[22] D. Martin, C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using brightness and texture. In *NIPS*, 2002. 2

[23] A. Mohan, C. Papageorgiou, and T. Poggio. Example-based object detection in images by components. *IEEE Trans. PAMI*, 23(4):349–361, 2001.

[24] S. Palmer. *Vision Science: Photons to Phenomenology*. MIT Press, 1999. 1

[25] X. Ren, A. C. Berg, and J. Malik. Recovering human body configurations using pairwise constraints between parts. In *ICCV*, volume 1, pages 824–831, 2005. 2

[26] X. Ren, C. Fowlkes, and J. Malik. Scale-invariant contour completion using conditional random fields. In *ICCV*, volume 2, pages 1214–1221, 2005. 2

[27] C. Schmid and A. Zisserman. Automatic line matching across views. In *CVPR*, 1997. 2

[28] K. Toyama and A. Blake. Probabilistic tracking in a metric space. In *ICCV*, 2001. 1