1 Review of RJMCMC

Starting with an initial state, RJMCMC proposes a new state $\mathbf{o}'$ from a proposal distribution $Q(\mathbf{o}'; \mathbf{o})$ that depends on the current state $\mathbf{o}$. The proposed state is probabilistically accepted as the next state in the Markov Chain according to the acceptance ratio

$$\alpha(\mathbf{o}, \mathbf{o}') = \min(1, \frac{\pi(\mathbf{o}')}{\pi(\mathbf{o})} \frac{Q(\mathbf{o}\ ;\ \mathbf{o}')}{Q(\mathbf{o}';\ \mathbf{o})} \times J_{F_{|\mathbf{o}| \to |\mathbf{o}'|}}), \tag{1}$$

where the target distribution π is the posterior distribution $P(\mathbf{o}|\mathbf{Z})$ (Eqn. 8) and J is the Jacobian determinant of a dimension matching function F [1], which for all of our move proposals simplifies to a constant value of one. Notice that we no longer need to worry about the intractable normalizing constant as it cancels out in the acceptance ratio. The algorithm is outlined in the table below.

Algorithm 1. RJMCMC

Input: $\mathbf{Z}$, W, ITERMAX
Output: $\mathbf{o}^*$
Initialization: $\mathbf{o} \leftarrow \mathbf{o}_0$
for $t = 1$ to ITERMAX
  choose a move c according to the distribution $p_c(\mathbf{o}^{t-1})$
  propose $\mathbf{o}'$ from the move-specific proposal $Q_c(\mathbf{o}^{t-1}; \cdot)$
  sample $u \sim \texttt{Uniform}(0, 1)$
  if $\log(u) < \log(\alpha(\mathbf{o}^{t-1}, \mathbf{o}'))$
    $\mathbf{o}^t \leftarrow \mathbf{o}'$
  else
    $\mathbf{o}^t \leftarrow \mathbf{o}^{t-1}$
  if $P(\mathbf{o}^t|\mathbf{Z}) > P(\mathbf{o}^*|\mathbf{Z})$
    $\mathbf{o}^* \leftarrow \mathbf{o}^t$

---

2 Additional Results for S2L1

- Ground-truth Annotation
  We compute the ground truth annotations by hand from four camera views, beginning by manually clicking on the head of each person in each view. For the people visible in more than one view, 3D head location is computed by triangulation of head viewing rays. For people visible in only one view, 3D head location is computed by intersecting the head viewing ray with a horizontal plane of average person height. For each computed (X,Y,Z) head location, we form a vertical 3D cylinder spanning from head position (X,Y,Z) to foot position (X,Y,0). The projection of this 3D cylinder into each 2D view gives us an initial set of 2D annotated bounding boxes. Small manual adjustments to the position of these 2D rectangular boxes are made as needed to better align the boxes to the people in each view.

- Performance Metrics
  We evaluated each algorithm based on the MODA (Multiple Object Detection Accuracy) and MODP (Multiple Object Detection Precision) metrics from the CLEAR evaluation framework [2]. For a frame t, let G_i^t denote the annotated box and D_i^t the detection box. A detection is counted as correct if the overlap ratio,

$$\mathrm{OL}_i^t = \frac{|G_i^t \cap D_i^t|}{|G_i^t \cup D_i^t|},$$

  is greater than some threshold τ. Letting N_G^t be the total number of annotated objects, N_m^t the number of correct detections, m_t the number of false negatives, and f_t the number of false positives, MODP measures the localization quality of the correct detections,

$$\mathrm{MODP}_t = \frac{\sum_{i=1}^{N_m^t} \mathrm{OL}_i^t}{N_m^t},$$

  while MODA is a detection accuracy measure taking into account both false negatives and false positives,

$$\mathrm{MODA}_t = 1 - \frac{m_t + f_t}{N_G^t}.$$

  For both metrics, larger values are better.

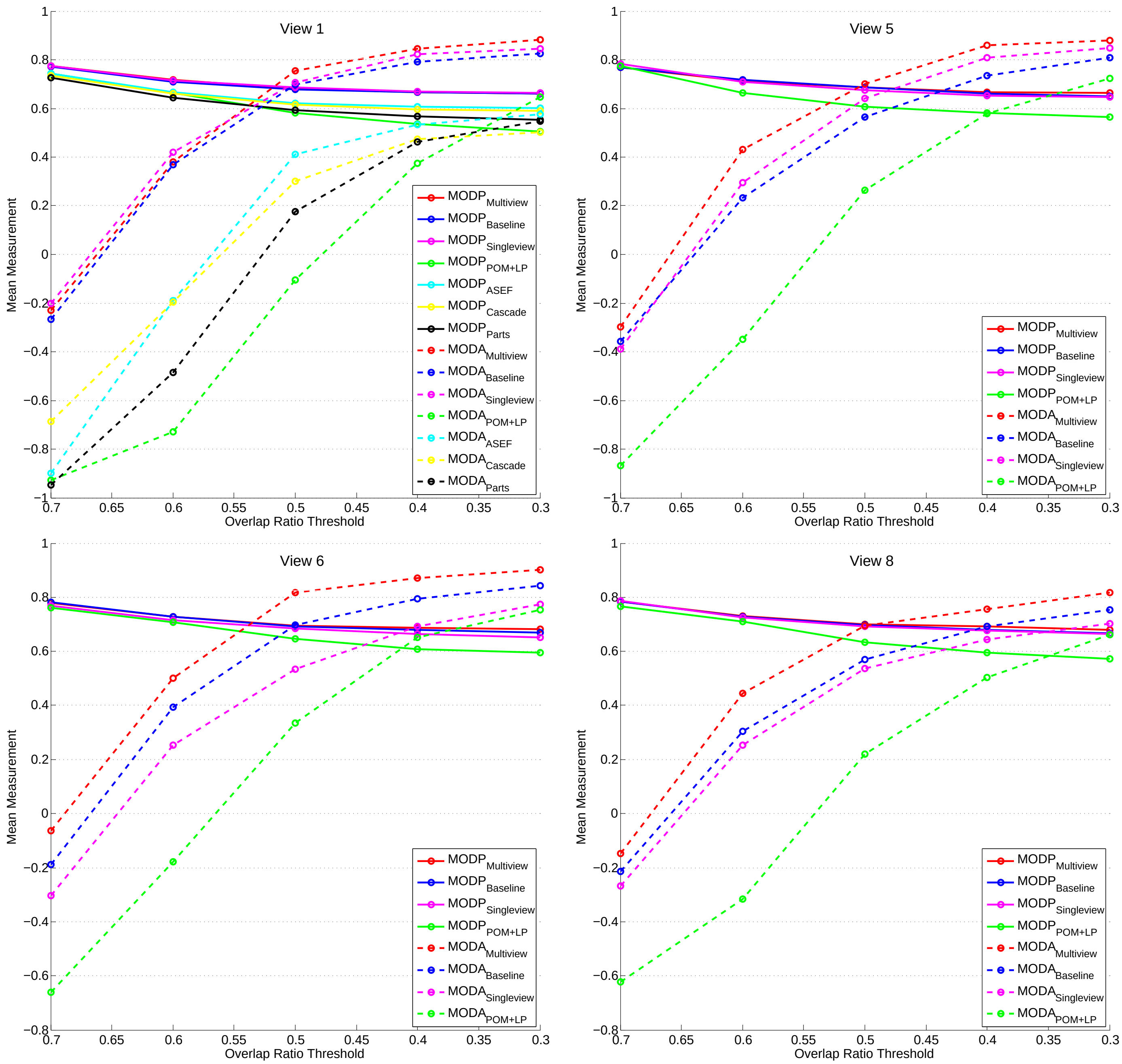

Figure 1: Evaluation results on S2L1 (MODP&MODA).

Table 1: Evaluation results on S2L1 (MODP&MODA).

Overlap	Method	View 1		View 5		View 6		View 8	
0.7	Multiview	0.7754	-0.2306	0.7702	**-0.2971**	0.7788	**-0.0626**	0.7837	**-0.1477**
	Baseline	0.7712	-0.2676	0.7693	-0.3574	**0.7810**	-0.1894	0.7834	-0.2146
	Singleview	**0.7758**	**-0.2019**	**0.7832**	-0.3896	0.7684	-0.3028	**0.7875**	-0.2667
	POM+LP	0.7441	-0.9269	0.7735	-0.8668	0.7604	-0.6620	0.7654	-0.6224
	ASEF	0.7431	-0.8993	-					
	Cascade	0.7336	-0.6878	-					
	Parts	0.7262	-0.9476	-					
0.6	Multiview	**0.7169**	0.3811	0.7123	**0.4298**	**0.7288**	**0.4998**	**0.7297**	**0.4447**
	Baseline	0.7101	0.3699	**0.7178**	0.2314	0.7270	0.3940	0.7276	0.3038
	Singleview	0.7135	**0.4198**	0.7092	0.2948	0.7162	0.2525	0.7266	0.2537
	POM+LP	0.6639	-0.7283	0.6652	-0.3487	0.7072	-0.1776	0.7095	-0.3165
	ASEF	0.6666	-0.1898	-					
	Cascade	0.6596	-0.1951	-					
	Parts	0.6449	-0.4859	-					
0.5	Multiview	0.6805	**0.7532**	**0.6872**	**0.6998**	**0.6953**	**0.8162**	**0.7004**	**0.6941**
	Baseline	0.6791	0.6988	0.6872	0.5660	0.6936	0.6967	0.6967	0.5702
	Singleview	**0.6863**	0.7052	0.6751	0.6415	0.6855	0.5333	0.6924	0.5357
	POM+LP	0.5806	-0.1037	0.6071	0.2630	0.6467	0.3354	0.6344	0.2188
	ASEF	0.6212	0.4116	-					
	Cascade	0.6150	0.3000	-					
	Parts	0.5927	0.1759	-					
0.4	Multiview	0.6673	**0.8467**	**0.6664**	**0.8602**	**0.6877**	**0.8707**	**0.6925**	**0.7559**
	Baseline	0.6659	0.7915	0.6597	0.7351	0.6802	0.7943	0.6799	0.6915
	Singleview	**0.6688**	0.8217	0.6528	0.8073	0.6654	0.6922	0.6764	0.6429
	POM+LP	0.5360	0.3756	0.5819	0.5798	0.6082	0.6514	0.5950	0.5028
	ASEF	0.6071	0.5343	-					
	Cascade	0.5952	0.4726	-					
	Parts	0.5672	0.4618	-					
0.3	Multiview	0.6603	**0.8830**	**0.6628**	**0.8784**	**0.6826**	**0.9027**	**0.6797**	**0.8181**
	Baseline	0.6597	0.8253	0.6497	0.8077	0.6702	0.8439	0.6679	0.7530
	Singleview	**0.6645**	0.8441	0.6455	0.8471	0.6517	0.7739	0.6654	0.7026
	POM+LP	0.5047	0.6460	0.5634	0.7223	0.5943	0.7535	0.5721	0.6621
	ASEF	0.6003	0.5772	-					
	Cascade	0.5903	0.5022	-					
	Parts	0.5539	0.5469	-					

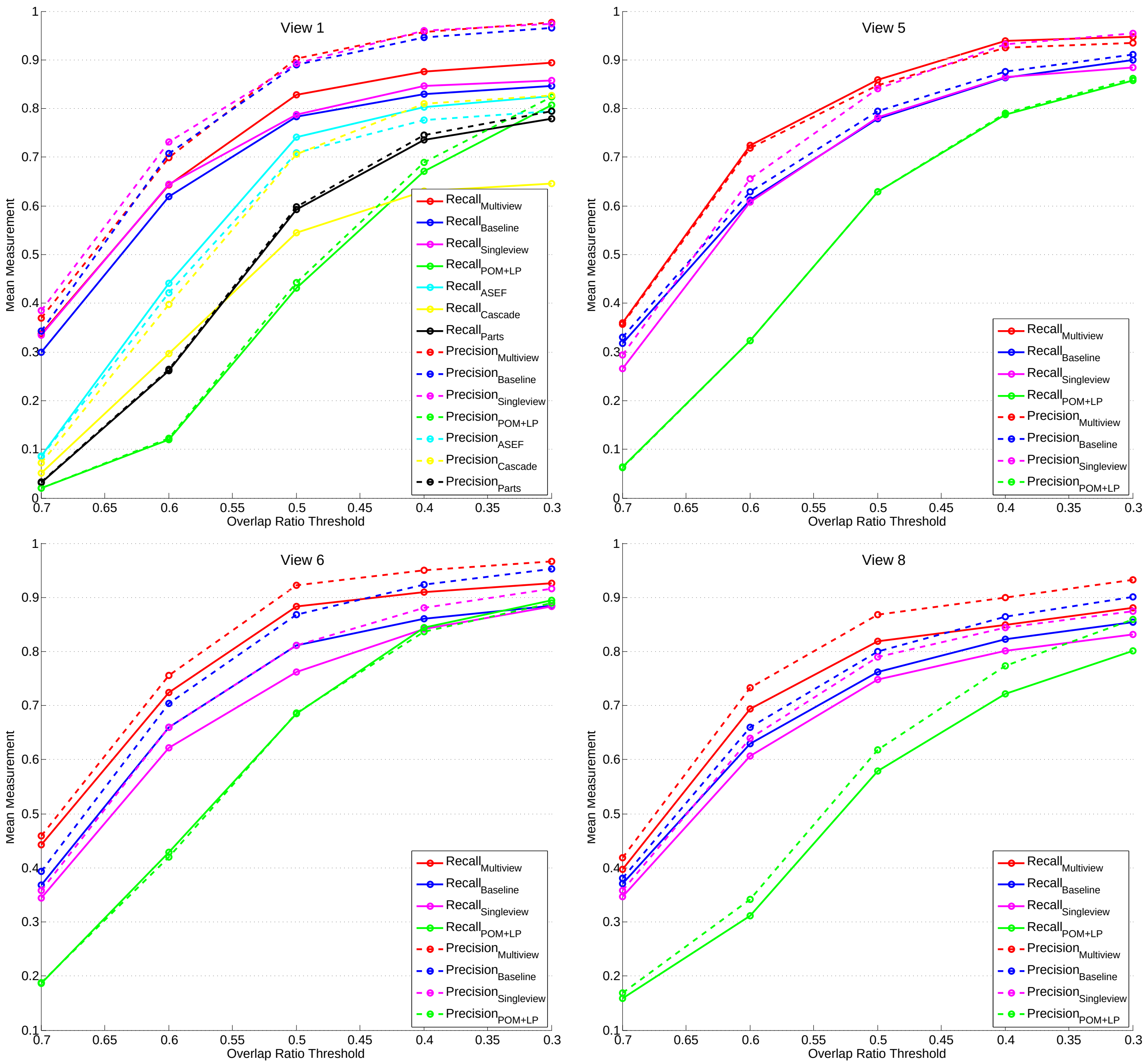

Figure 2: Evaluation results on S2L1 (Recall&Precision).

Table 2: Evaluation results on S1L1 (Recall&Precision).

Overlap	Method	View 1		View 5		View 6		View 8	
0.7	Multiview	**0.3369**	0.3697	**0.3603**	**0.3563**	**0.4432**	**0.4588**	**0.3976**	**0.4192**
	Baseline	0.2998	0.3423	0.3172	0.3304	0.3681	0.3935	0.3702	0.3807
	Singleview	0.3342	**0.3845**	0.2660	0.2944	0.3446	0.3583	0.3468	0.3584
	POM+LP	0.0201	0.0203	0.0638	0.0621	0.1869	0.1875	0.1585	0.1685
	ASEF	0.0865	0.0853	-					
	Cascade	0.0509	0.0726	-					
	Parts	0.0311	0.0323	-					
0.6	Multiview	0.6428	0.6987	**0.7238**	**0.7192**	**0.7244**	**0.7560**	**0.6938**	**0.7332**
	Baseline	0.6185	0.7074	0.6116	0.6294	0.6598	0.7045	0.6294	0.6603
	Singleview	**0.6450**	**0.7308**	0.6082	0.6556	0.6223	0.6594	0.6071	0.6390
	POM+LP	0.1194	0.1228	0.3228	0.3235	0.4292	0.4202	0.3114	0.3423
	ASEF	0.4412	0.4213	-					
	Cascade	0.2973	0.3974	-					
	Parts	0.2620	0.2644	-					
0.5	Multiview	**0.8288**	**0.9022**	**0.8588**	**0.8481**	**0.8826**	**0.9223**	**0.8185**	**0.8677**
	Baseline	0.7830	0.8903	0.7789	0.7941	0.8111	0.8683	0.7626	0.8004
	Singleview	0.7878	0.8934	0.7815	0.8414	0.7627	0.8115	0.7480	0.7893
	POM+LP	0.4317	0.4429	0.6287	0.6288	0.6857	0.6860	0.5791	0.6185
	ASEF	0.7419	0.7085	-					
	Cascade	0.5449	0.7061	-					
	Parts	0.5929	0.5977	-					
0.4	Multiview	**0.8756**	0.9566	**0.9390**	0.9250	**0.9098**	**0.9506**	**0.8494**	**0.8991**
	Baseline	0.8293	0.9455	0.8635	0.8757	0.8599	0.9235	0.8232	0.8641
	Singleview	0.8460	**0.9607**	0.8644	**0.9313**	0.8421	0.8803	0.8016	0.8443
	POM+LP	0.6713	0.6888	0.7871	0.7898	0.8436	0.8370	0.7211	0.7737
	ASEF	0.8033	0.7764	-					
	Cascade	0.6311	0.8102	-					
	Parts	0.7358	0.7448	-					
0.3	Multiview	**0.8937**	**0.9764**	**0.9481**	0.9343	**0.9259**	**0.9667**	**0.8805**	**0.9326**
	Baseline	0.8462	0.9659	0.8998	0.9109	0.8847	0.9521	0.8540	0.9009
	Singleview	0.8572	0.9740	0.8843	**0.9539**	0.8829	0.9164	0.8315	0.8742
	POM+LP	0.8065	0.8242	0.8583	0.8618	0.8947	0.8894	0.8007	0.8594
	ASEF	0.8247	0.7941	-					
	Cascade	0.6459	0.8269	-					
	Parts	0.7784	0.7941	-					

References

[1] Green, P.: Reversible jump Markov chain Monte-Carlo computation and Bayesian model determination. Biometrika **82** (1995) 711–732

[2] Kasturi, R., Goldgof, D., Soundararajan, P., Manohar, V., Garofolo, J., Boonstra, M., Korzhova, V., , Zhang, J.: Framework for performance evaluation of face, text, and vehicle detection and tracking in video: Data, metrics, and protocol. IEEE Transactions on Pattern Analysis and Machine Intelligence **31** (2009) 319–336