ECCV 2012
European Conference on Computer Vision 2012
October 7-13, 2012, Firenze , Italy

Reviews For Paper

Track	Main Conference
Paper ID	1485
Title	Predicting Structured Geometry From Single And Multiple Views

Masked Reviewer ID: Assigned_Reviewer_1
Review:

Question	
Provide a brief description of the contribution and explain your overall assessment by listing positive and negative points.	The paper proposes a new indoor manhattan structure estimation algorithm from single/multiple views. The paper proposed an interesting algorithm with compelling results. However, it misses the most important paper to be cited (Manhattan Scene Understanding Using Monocular, Stereo and 3D Features, Alex Flint, 2011). See my comments below for the details.
Novelty	Incremental
Explain your novelty rating and provide relevant citations if you find it incremental or done before.	Very similar thing has been done by [Flint et al., 2011], which is not cited.
Technical correctness. Is the manuscript technically sound?	Yes
Technical comments: Explain your assessment of technical correctness.	I did not see a flaw.
Experimental Validation.	Insufficient
Explain your assessment of experimental validation.	I appreciate all the evaluations against [4]. However, real comparison must be done against [Flint et al., 2011].
All other comments including clarity of presentation, any	As noted above, the paper misses the most important reference to be cited [Flint et al., 2011]. If [Flint et al., 2011] did not exist, I would have no complaints about this paper. However, the paper exists, and I am very surprised that [Flint et al., 2011] is not cited in this paper. The problem is that an algorithm and contributions of the paper are very similar to [Flint et al., 2011]. It is also not clear on the results if this paper is doing better than [Flint et al., 2011]. After looking at and comparing results qualitative by my eyes, I could not see clear improvements by this

missed references, supplemental material.	paper over [Flint et al., 2011], which unfortunately means that this paper does not add new values to the community when published. Authors should include a reference to [Flint et al., 2011], rewrite the paper to clarify the difference from [Flint et al., 2011], and shows that it works better.
Overall Rating.	Probable Reject
Adherence to submission guidelines (paper length, anonymity, double submission, supplementary material). If a violation is suspected, contact the program chairs (eccv-2012-program-chairs@googlegroups.com) immediately. You must continue to review the paper as if guidelines were followed until the PCs say otherwise.	Adheres to guidelines

Masked Reviewer ID: Assigned_Reviewer_2

Review:

Question	
Provide a brief description of the contribution and explain your overall assessment by listing positive and negative points.	The authors propose structured prediction to learn the weights in the MAP energy for infering Manhattan-geometry from single and multiple views. Positive points: + Formulates learning the weights for Manhattan-geometry inference in a principled way Negative aspects: - Novelty is incremental - Although the material is not complicated by itself, the description makes everything difficult to comprehend - Many minor mistakes and errors indicate, that the submission was finished in a rush
Novelty	Incremental
Explain your novelty rating and provide relevant citations if you find it incremental or done before.	Training the importance weighs of features in a structured prediction task using SSVMs is not highly innovative IMO. This submission presents little beyond setting up the problem, and describing the particular SSVM training procedure.

Technical correctness. Is the manuscript technically sound?	Has minor problems
Technical comments: Explain your assessment of technical correctness.	The presented material is not complicated at all, but due to several glitches I rate the technical part to have minor (but easy to fix) errors. First, I got the very strong impression that the authors have an incorrect citation for [4], since from the description they really mean Flint's ICCV 2011 paper. Since the authors heavily build on [4], such mistake is very confusing to the reader and should really not happen. Second, several entities should be better defined (or defined at all). o_i in Eq.1 is only vagely defined, and in Eq.3 the authors should clarify if they really mean distance (along a camera ray) or depth (distance to the image plane). nu_theta and nu_i are not defined at all. What does the index i denote in Eq.7ff? One can only guess it refers to different feature modalities. The double use of n (denoting the number of walls in Eq.4 and something else in Eq.7) does not help for easy understanding. The authors use one page to show, that the MAP energy can be extended to a structured prediction task by adding weights to the potentials. That seems so obvious to me and I don't see the contribution of "place this model into a statistically rigorous learning framework" (L209).
Experimental Validation.	Sufficient
Explain your assessment of experimental validation.	I cannot complain about the experiments. The authors trained and evaluated on two meaningful loss functions, depth-based and labeling-based error. I don't know what to think about the scatter plot in Fig.5. One would expect that training using depth loss or using labeling loss would make little difference, since the Manhattan assumption very strongly links wall orientation and depth. Maybe the plot in Fig.5 is an artefact of limited training data.
All other comments including clarity of presentation, any missed references, supplemental material.	.
Overall Rating.	Borderline Poster/Reject
Adherence to submission guidelines (paper length, anonymity, double submission, supplementary material). If a violation is suspected, contact the program chairs (eccv-2012-program-chairs@googlegroups.com) immediately. You must continue to review the paper	Adheres to guidelines

as if guidelines were followed until the PCs say otherwise.	

Masked Reviewer ID: Assigned_Reviewer_3

Review:

Question	
Provide a brief description of the contribution and explain your overall assessment by listing positive and negative points.	This paper formulates single-view or multiple-view reconstruction as a learning problem. Scenes are modeled using the indoor Manhattan world assumptions and structured prediction is used to learn predictors that map features (mostly photometric features) to reconstructions. The work is interesting in a continuation of previous work on the problem [4,5], where efficient inference methods were developed for this problem. This paper focuses on learning all model parameters from training data using structured prediction. However, the paper is not self-contained, and difficult to read in its current version. There is emphasis on the theoritical aspect of the learning problem but key details about features, important notations is missing (see comments below). Experimental results are presented in both the single-view and multiple-view settings for multiple datasets. However, the evaluation methodology is not very convincing (see comments). The training data provided from [4] does not seem to be very high quality and this makes it difficult to interpret the value of the proposed learning framework in the context of the experiments. I disagree with the claim in lines 570 - 572 that this method outperforms the state of the art in multi-view reconstruction. Overall the paper has interesting ideas but this version of the paper has multiple shortcomings explained below.
Novelty	Moderately original
Explain your novelty rating and provide relevant citations if you find it incremental or done before.	Learning the parameters of this model using structured prediction is a novel contribution as far as I know. However since the core steps of the structured SVM algorithm [6] is solved using the known inference algorithm of [5], the novelty is really in applying structured SVMs to this problem. The 3d scene model and features used in the framework seem to be similar to

	what is used in prior work.
Technical correctness. Is the manuscript technically sound?	Has minor problems
Technical comments: Explain your assessment of technical correctness.	Some notation in Section 3.2 is not explained. - Equation 4: what is "+ O(1)" ? what is \bold{n} ? The phrase "a vector containing the number of walls ..." is unclear. Equation 6: v_\theta(x,s_x) is not defined. using "n" once with bold-face and once without is confusing. Section 3.3: "\Delta_{labeling} is omitted for space constraints. This should have been in the supplementary material at least. Equation 22: \gamma is not defined. What is it ?
Experimental Validation.	Insufficient
Explain your assessment of experimental validation.	Several experiments are reported using datasets from [4]. Although, visually the results look better than [5] in some cases, the multi-view reconstructions often have gross errors. Are these using any multi-view cues at all ? Are the corresponding weights very small ? Errors in results like Fig7-row1-col4, are strange given that multiview cues (photoconsistency etc.) are being used. I noticed in some visual examples in the paper and supplementary that the ground truth labeling has errors. For example in Figure 7, images in row 4, columns 3-5 have incorrectly labeled pixels. If the quality of the ground truth is low, can one really expect the method to work well ? Using ground truth depth using Kinect or even a high quality multiview reconstruction system could be a lot more accurate. Are there any images in the test sequences of scenes not captured in the training set ? or are all test images taken from the same scenes but different viewpoints ? The evaluation methodology explained in Sec 5 seems inappropriate to me. By ignoring the surface orientation of the walls, the labeling error metric becomes quite unreliable I think.
	The paper is not self-contained making it difficult to read. (1) several key details about the indoor Manhattan model are unclear.

All other comments including clarity of presentation, any missed references, supplemental material.	lines 160 onwards .. [.. do not have space for a full discussion ..] - but this makes the paper difficult to understand. - how is metric depth computed from a seam S ? what are the assumptions ? - can depth be computed when the floor/ceiling is not visible ? how ? - it appears to me that doors cannot be represented in this model ? these and other limitations should at least be briefly mentioned. - line 166: regarding "orientation of 3d surface". Is a(x,y;S) a binary label indicating either red or green walls ? (2) What are the concrete image features used ? - line 237-239 and Figure 2 is all I found. - Why is the dimensionality of photo-consistency features = 4 ? point cloud = 2 ? - what are 'line sweeps' features ? Typos: - Reference [20] is repeated .. - Reference [21] misspelt. - line 42 "part part" - line 572 "future work will" .. - Figure 6 is using other features like 'corner penalty/occlusion penalty/stereo offset +5 what are these ?
Overall Rating.	Probable Reject
Adherence to submission guidelines (paper length, anonymity, double submission, supplementary material). If a violation is suspected, contact the program chairs (eccv-2012-program-chairs@googlegroups.com) immediately. You must continue to review the paper as if guidelines were followed until the PCs say otherwise.	Adheres to guidelines