CVPR 2016
IEEE Conference on Computer Vision and Pattern Recognition 2016
Las Vegas, USA

Reviews For Paper

Paper ID	2049
Title	3D Reconstruction from Multimodal and Coaxial Camera Rigs Using Image Correspondences Derived from Perceived Motion

Masked Reviewer ID: Assigned_Reviewer_17

Review:

Question	
Paper Summary	This paper presents a new stereo reconstruction tailored to the case of coaxial cameras, where one camera is virtually "behind" the other one (by using a mirror and beamsplitter). In order to use cameras with different spectral characteristics, the paper proposes to compute the optical flow and use the flow itself for the disparity matching. Clearly, this should allow very differents image modalities to be matched, as long as there is scene motion to be measured.
Paper Strengths. Please discuss the positive aspects of the paper. Be sure to comment on the paper's novelty, technical correctness, clarity and experimental evaluation. Notice that different papers may need different levels of evaluation: a theoretical paper may need no experiments, while a paper presenting a new approach to a known problem may require thorough comparisons to existing methods. Also, please make sure to justify your comments in great detail. For example, if you think the paper is	The idea of using a coaxial setup for stereo combining the optical flow for matching is innovative, in my opinion. The derivation which links disparity, stereo geometry and optical flow is interesting and might find other use for non coaxial cameras. The experimental results, especially the disparity maps, suggests that the method works. The paper is clearly written.

novel, not only say so, but also explain in detail why you think this is the case.	
Paper Weaknesses. Please discuss the negative aspects of the paper: lack of novelty or clarity, technical errors, insufficient experimental evaluation, etc. Justify your comments in great detail. If you think the paper is not novel, explain why and give a reference to prior work. Keep in mind that novelty can take a number of forms; a paper may be novel in terms of the method, the problem, the theory, analysis for an existing problem, or the empirical evaluation. If you think there is an error in the paper, explain in detail why it is an error. If you think the experimental evaluation is insufficient, remember that theoretical results/ideas are essential to CVPR and that a theoretical paper need not have experiments. It is *not* okay to reject a paper because it did not outperform other	-- experimental results -- The synthetic results provided (Figs 4,5,6,7,8,9,10,11) take up a lot of space, given that they do not provide very much information. A table could have summerized all those results adequately, in my opinon, and would have provided more space for results on real images. The scene used for synthetic results is not very well described ("10m to 20m depth range" L.503, "smooth scene without occlusion" L.510, "large 8m discontinuity" L.549). It would be very useful to provide an actual image of the scene, with its true depth map, and possibly the recovered depth map. The use of RMS error measure for disparity is often of limited usefulness, because it does not provide the variation of errors, or its distribution in the image. It could be useful to provide a depth map or error map to supplements those RMS results. --- real results --- Since the goal is to match multimodal cameras, it would have been very informative to see the stereo pair where one image is RGB and the other is IR, so we can appreciate the difference. Also, for the stereo rig as well as the coaxial rig, it would have been useful to see the illustrated optical flow obtained (by the Borx & Malik algorithm) and a sample disparity map. This disparity map would help understand the results of figure 17 and 18. Also, it seems that the "stereo rig" and "Coaxial rig" were not run with the same scenes (one is with Fig 12, one with Fig 15). Why not run "stereo rig RGB+IR", "stereo rig RGB+RGB", "coaxial rig RGB+IR" and "coaxial RGB+RGB" on both scenes, and obtain 8 results which are easy to compare? Currently, the paper only provides "stereo RGB+IR" on scene 1, and "coaxial RGB+RGB" on scene 2.

existing algorithms, especially if the theory is novel and interesting. It is not reasonable to ask for comparisons with unpublished, non peer reviewed papers (e.g. ArXiv) or papers published after the CVPR'16 deadline.	
Preliminary Rating. This rating indicates to the area chair, to other reviewers, and to the authors, your current opinion on the paper. Please use 'Borderline' only if the author rebuttal and/or discussion might sway you in either direction.	Strong Reject
Preliminary Evaluation. Please explain to the AC, your fellow reviewers, and the authors your current opinion on the paper. This explanation may include how you weigh the importance of the various strengths and weaknesses you described above in Q1-Q3. Please summarize the key things you would like the authors to include in their rebuttals to facilitate your decision making. There is no need to summarize the	Even if the problem of multimodal stereo is interesting, and the paper propose interesting ideas to tackle it (especially matching the optical flow), I strongly feel that the experimental results do not provide adequate information to assess the performance and quality of the method.

paper.	
New exciting ideas. CVPR'16 would like to draw attention to papers that explore highly innovative ideas, novel problems, and/or paradigm shifts in conventional theory and practice. Such papers may not be "complete" in the "traditional" manner in the sense that it may not be possible to have experimental results comparing other related efforts or that they may not have large, publicly available data sets to be used for performance comparison. However, we expect these papers to be visionary by nature. Should this paper be considered under "new exciting ideas"?	No
Reproducibility. Could the work be reproduced from the information in the paper? Are all important algorithmic or system details discussed adequately?	The paper is clearly written and reproducible, although the experimental results would be hard to reproduce because of the lack of sample images in the paper. Obviously, given the specialization of the experimental setup, it would be useful to provide the stereo and coaxial image sequences themselves to make it easy to reproduce the proposed method, as well as eventually compare with other algorithms.
Confidence. Select: "Very Confident" to stress that you	

are absolutely sure about your conclusions (e.g., you are an expert who works in the paper's area), "Confident" to stress that you are mostly sure about your conclusions (e.g., you are not an expert but can distinguish good work from bad work in that area), and "Not Confident" to stress that that you feel some doubt about your conclusions. In the latter case, please provide details as confidential comments to PC/AC chairs (point 7.)	Very Confident

Masked Reviewer ID: Assigned_Reviewer_25

Review:

Question	
Paper Summary	The paper deals with 3D reconstruction from an image pair sequence acquired using a multi-modal or a coaxial camera rig. The main idea of the paper is two avoid performing disparity estimation between simultaneously acquired pairs of images which is problematic in the case of a multimodal rig; instead the approach performs optical flow estimation separately for each camera and then recovers the disparity by optimising for a disparity map (or equivalently a depth map) which enforces coherence of the flow maps. A variational approach is used to solve the problem, which separate implementations provided for the multi-modal and the coaxial setups. The evaluation is performed on a small number of synthetic scenes and one real scene.
Paper Strengths. Please discuss the positive aspects of the paper. Be sure to comment on the paper's novelty, technical correctness, clarity and experimental	

evaluation. Notice that different papers may need different levels of evaluation: a theoretical paper may need no experiments, while a paper presenting a new approach to a known problem may require thorough comparisons to existing methods. Also, please make sure to justify your comments in great detail. For example, if you think the paper is novel, not only say so, but also explain in detail why you think this is the case.	- The idea of performing disparity estimation from optical flow input rather than image correspondences seems original. Stereoscopic scene flow methods such as those proposed by Huguet & Devernay "A variational method for scene flow estimation from stereo sequences, ICCV 2007 & Wedel et al "Stereoscopic Scene Flow Computation for 3D Motion Understanding", IJCV2011 typically simultaneously enforce both optical flow and disparity constraint. Technically, the main novelty is the introduction of the constraints in equations (1) and (6) - though I did not check the correctness as derivations have not been provided. The formulation of the energies are based on a data term minimising the squared error defined by these constraints and a quadratic regulariser, which is standard.
Paper Weaknesses. Please discuss the negative aspects of the paper: lack of novelty or clarity, technical errors, insufficient experimental evaluation, etc. Justify your comments in great detail. If you think the paper is not novel, explain why and give a reference to prior work. Keep in mind that novelty can take a number of forms; a paper may be novel in terms of the method, the problem, the theory, analysis for an existing problem, or the	- Clarity: the method is not very clearly discussed with several derivations omitted. For example, the derivation of equations (1) and (6) are important as this is the basis of the method. It is therefore important that these are provided in appendix or, if space does not permit, in supplementary material. In section 4.1, for completeness the Euler-Lagrange could also be provided in supplementary material to facilitate reproduction of the results (and also to check correctness). Line 448 I could not follow why there is a pixel pair on each epipolar line satisfying x_l=-x_r (maybe I missed something). - Technical correctness: the convergence of the method is unclear to me. From equations (2), (7) and (8) it is clear that the optimised energies are very non-linear, so it seems very likely that a gradient descent method will get stuck in a local minimum unless initialised very close to the solution. There are several statements made in the paper which seem to indicate some convergence difficulties. First from section 4.2 it appears that the method works best if the first image pair contains translation purely along X and Z; this seems to be a severe practical limitation. Second the stopping criteria are rather vaguely described. They are based on the quality of the flow fields

empirical evaluation. If you think there is an error in the paper, explain in detail why it is an error. If you think the experimental evaluation is insufficient, remember that theoretical results/ideas are essential to CVPR and that a theoretical paper need not have experiments. It is *not* okay to reject a paper because it did not outperform other existing algorithms, especially if the theory is novel and interesting. It is not reasonable to ask for comparisons with unpublished, non peer reviewed papers (e.g. ArXiv) or papers published after the CVPR'16 deadline.	but how are these assessed and how do we draw the line between good and bad flow fields? The stopping criteria in each seem based on heuristics. Does the method diverge if not stopped? - Application/scope: the general idea seems interesting from an academic point of view, but it is not clear why in practice such type of rig is useful. It would have been useful to discuss situation where this type of rig is required to better motivate the work - Evaluation: The evaluation is limited in terms of the number of scenes considered and also due to lack of comparison against other methods. A few different synthetic scenes have been considered and tested under different camera motion speed, however it would be useful to show some information about the actual scenes and some qualitative results comparing the results obtained against the ground truth. The technique should be evaluated on a larger number of real scenes (only one considered). It is quite hard to assess the quality of the results as there is no comparison and limited information about the camera displacement. It is true that classical methods performing disparity estimation are likely to fail with the multi-modal rig, making a comparison difficult, but I do not seem any reason why the comparison cannot be performed in the case of the coaxial camera rig. It seems important to compare against monocular camera tracking methods such as SfM and Visual SLAM as these have reach a high quality. See for example: Newcombe& Davison, Live Dense Reconstruction with a Single Moving Camera. CVPR10
Preliminary Rating. This rating indicates to the area chair, to other reviewers, and to the authors, your current opinion on the paper. Please use 'Borderline' only if the author rebuttal and/or discussion might sway you in either direction.	Weak Reject
Preliminary Evaluation. Please explain to the AC, your fellow	

reviewers, and the authors your current opinion on the paper. This explanation may include how you weigh the importance of the various strengths and weaknesses you described above in Q1-Q3. Please summarize the key things you would like the authors to include in their rebuttals to facilitate your decision making. There is no need to summarize the paper.	The general idea of the approach seems novel and may have some potential but I feel that at the moment the method relies too heavily on heuristics, there are some clarity issues with some derivations missing and also the evaluation is currently limited due to lack of comparison against other methods including SfM/SLAM. These are the three aspects which need to be addressed in the rebuttal.
New exciting ideas. CVPR'16 would like to draw attention to papers that explore highly innovative ideas, novel problems, and/or paradigm shifts in conventional theory and practice. Such papers may not be "complete" in the "traditional" manner in the sense that it may not be possible to have experimental results comparing other related efforts or that they may not have large, publicly available data sets to be used for performance comparison. However, we expect these papers to be	No

visionary by nature. Should this paper be considered under "new exciting ideas"?	
Reproducibility. Could the work be reproduced from the information in the paper? Are all important algorithmic or system details discussed adequately?	It would be difficult to reproduce the results due to some missing details. In particular the stopping criteria that were used in the evaluation do not seem to be specified and the Euler-Lagrange equations are not provided.
Confidence. Select: "Very Confident" to stress that you are absolutely sure about your conclusions (e.g., you are an expert who works in the paper's area), "Confident" to stress that you are mostly sure about your conclusions (e.g., you are not an expert but can distinguish good work from bad work in that area), and "Not Confident" to stress that that you feel some doubt about your conclusions. In the latter case, please provide details as confidential comments to PC/AC chairs (point 7.)	Confident

Masked Reviewer ID: Assigned_Reviewer_26

Review:

Question	
	The paper addresses an interesting topic of cross modality image

Paper Summary	correspondence estimation for 3D reconstruction and 3D reconstruction for coaxial cameras. In particular, the paper introduces two approaches to leverage optical flow to perform 3D structure estimation for image pairs of cameras that move with respect to the scene. Specifically, the authors assume that optical flow is available for each image (estimated from the scene motion over time captured by each of the cameras). The optical flow is then used within the two proposed variational approaches to determine the correspondence field between the image pairs, which can subsequently be turned into a depth map. The approach is applied to a multimodal binocular stereo system (RGB, IR) and a coaxial camera and is evaluated for each on a synthetic and a real scene.
Paper Strengths. Please discuss the positive aspects of the paper. Be sure to comment on the paper's novelty, technical correctness, clarity and experimental evaluation. Notice that different papers may need different levels of evaluation: a theoretical paper may need no experiments, while a paper presenting a new approach to a known problem may require thorough comparisons to existing methods. Also, please make sure to justify your comments in great detail. For example, if you think the paper is novel, not only say so, but also explain in detail why you think this is the case.	+ the paper introduces a novel approach for estimating correspondences not directly leveraging visual correspondence but image sequence motion + the method is technically sound
Paper Weaknesses. Please discuss the negative aspects of the paper: lack of novelty or clarity, technical errors, insufficient	

experimental evaluation, etc. Justify your comments in great detail. If you think the paper is not novel, explain why and give a reference to prior work. Keep in mind that novelty can take a number of forms; a paper may be novel in terms of the method, the problem, the theory, analysis for an existing problem, or the empirical evaluation. If you think there is an error in the paper, explain in detail why it is an error. If you think the experimental evaluation is insufficient, remember that theoretical results/ideas are essential to CVPR and that a theoretical paper need not have experiments. It is *not* okay to reject a paper because it did not outperform other existing algorithms, especially if the theory is novel and interesting. It is not reasonable to ask for comparisons with unpublished, non peer reviewed papers (e.g. ArXiv) or papers published after the CVPR'16	- the weak part of the paper is the evaluation on real data. While I realize that this maybe some effort I think it is doable for the authors. In particular it would be good to evaluate the approach on more realistic scenes than the ones used (Fig. 12, Fig. 15). Without a convincing evaluation on real data, for example also the suggested endoscopic data, it is hard more me to judge the value of the contribution. - the authors observer in 599-603 that there could be a limitation on how large the Z displacement can be for their method. However, they do not present any discussion or analysis for this limitation, which would have been nice to provide. - there seems to be a slight disconnect in the introduction. In line 161 you motivate with correspondence finding techniques but then later in line 165-166 (referring to it) you talk about it not producing a sufficient disparity for triangulation. These are actually two distinct problems as correspondence finding establishes the correspondence in the image and triangulation takes a correspondence to estimate the observed 3D point. The confusion of the two seems to reoccur in other parts of the paper as well. It would be good to clearly distinguish between the two as it easily confuses the reader. - it would be good to show both images for the scenes in Fig. 12 and Fig. 15 as that would make it easier to judge the difficulty the method is facing. - it seems to me that the b in Eq. (3), (4), (9) and (10) is not the baseline, which was defined as b in lines 296 and 349+1 (note not equal to 350 as there are access lines)

deadline.	
Preliminary Rating. This rating indicates to the area chair, to other reviewers, and to the authors, your current opinion on the paper. Please use 'Borderline' only if the author rebuttal and/or discussion might sway you in either direction.	Borderline
Preliminary Evaluation. Please explain to the AC, your fellow reviewers, and the authors your current opinion on the paper. This explanation may include how you weigh the importance of the various strengths and weaknesses you described above in Q1-Q3. Please summarize the key things you would like the authors to include in their rebuttals to facilitate your decision making. There is no need to summarize the paper.	I'm on the edge about the recommendation of the paper. On one hand it has a clear contribution and a sound theory behind it but on the other hand due to the limited evaluation it is hard to value the relevance and robustness of the method. Since I don't see the paper as a theory paper the evaluation is important.
New exciting ideas. CVPR'16 would like to draw attention to papers that explore highly innovative ideas, novel problems, and/or paradigm shifts in conventional theory and	

practice. Such papers may not be “complete” in the “traditional” manner in the sense that it may not be possible to have experimental results comparing other related efforts or that they may not have large, publicly available data sets to be used for performance comparison. However, we expect these papers to be visionary by nature. Should this paper be considered under “new exciting ideas”?	No
Reproducibility. Could the work be reproduced from the information in the paper? Are all important algorithmic or system details discussed adequately?	The paper provides sufficient detail to be reproduced.
Confidence. Select: "Very Confident" to stress that you are absolutely sure about your conclusions (e.g., you are an expert who works in the paper's area), "Confident" to stress that you are mostly sure about your conclusions (e.g., you are not an expert but can distinguish good	Confident

work from bad work in that area), and "Not Confident" to stress that that you feel some doubt about your conclusions. In the latter case, please provide details as confidential comments to PC/AC chairs (point 7.)	

Masked Reviewer ID: Assigned_Reviewer_8

Review:

Question	
Paper Summary	This paper proposes a method for 3D reconstruction from two cameras of different modalities, where the intensities cannot be matched, using only the optical flow computed from both images. Two camera configurations are studied: the stereo case and the coaxial case. Of course, this method is supposed to work only if either the scene or the camera has motion.
Paper Strengths. Please discuss the positive aspects of the paper. Be sure to comment on the paper's novelty, technical correctness, clarity and experimental evaluation. Notice that different papers may need different levels of evaluation: a theoretical paper may need no experiments, while a paper presenting a new approach to a known problem may require thorough comparisons to existing methods. Also, please make sure to justify your comments in	The paper presents a rather new stereo matching approach, where the optical flows in both images are matched rather than the intensity. The equations giving the optical flow correspondences seem correct, although I do not understand why the study is limited to two camera configurations.

great detail. For example, if you think the paper is novel, not only say so, but also explain in detail why you think this is the case.	
Paper Weaknesses. Please discuss the negative aspects of the paper: lack of novelty or clarity, technical errors, insufficient experimental evaluation, etc. Justify your comments in great detail. If you think the paper is not novel, explain why and give a reference to prior work. Keep in mind that novelty can take a number of forms; a paper may be novel in terms of the method, the problem, the theory, analysis for an existing problem, or the empirical evaluation. If you think there is an error in the paper, explain in detail why it is an error. If you think the experimental evaluation is insufficient, remember that theoretical results/ideas are essential to CVPR	** Lack of novelty: Matching optical flows in stereo pairs is also done by scene flow methods, although most of these either also match intensities, or use another reconstruction method for the initial geometry. The original scene flow paper (Vedula, Baker, Kanade, 1999) describes in section 4.3 a method to reconstruct entirely from optical flow, but I think it requires at least 3 cameras, or maybe even 4 non-coplanar cameras (note that it also explicitly takes into account the aperture problem) ** Technical errors & lack of clarity: The biggest issue with the present paper is the part on initialization, which is actually the hardest part, since once you have the initial surface, the optical flow in both images can be used to reconstruct the scene flow, and thus scene geometry. However, the section on initialization is very short and very vague. I do not understand lines 446-449, and I do not see where equation (12) comes from. It also seems from the text that the depth can only be computed at the principal point? Another very big issue is that different geometries can actually produce the same optical flow, and I do not understand how the algorithm behaves in this case. Take the simple example of a fronto-parallel plane in the stereo configuration. A translation of this plane in the Z direction will produce the same optical flow, whatever the actual distance of the plane is. Similarly, a translation in the XY direction will produce the same uniform optical flow in both images, and I do not understand how it could be possible to recover depth from that. It is even probable that it is not possible to recover depth from two optical flows for any plan which is parallel to the line joining the two optical centers, be it in the stereo or in the coaxial configuration. That makes a lot of planes! If there are 3 cameras, only one plane orientation is degenerate, and with 4 non-coplanar cameras there should be no degenerate configuration. In section 4.1, the numerical method used to solve the problem should at least be mentioned (the explanations are very terse). ** Experimental evaluation: That method should at least be compared to a state-of-the-art multimodal stereo matching method. I do not see the point in having the "synthetic optical flow field" experiment, where the optical flow is the actual projection of the motion. Of course the errors are small in this case, since the input data is almost exact. For that experiment, the scene is described as "smooth", but is not shown.

and that a theoretical paper need not have experiments. It is *not* okay to reject a paper because it did not outperform other existing algorithms, especially if the theory is novel and interesting. It is not reasonable to ask for comparisons with unpublished, non peer reviewed papers (e.g. ArXiv) or papers published after the CVPR'16 deadline.	Experiments could be made on real optical flow computed on a synthetic scene instead (there are lots of datasets around for that). The results on real flow computed on real scenes seem rather bad. a RMS disparity error of several pixels (fig 13) is not satisfactory. Nothing is said about how the ground truth data is built. In the coaxial case, an "RMS disparity error [of] less than 1%" is mentioned, but 1% of what? ** Typos, etc: The format is probably wrong, because the last two lines have no numbers. I guess [19] should be Baker, Scharstein et al, IJCV 2011. The year of [18] is 2002, not 2001. l.249++: "has been cost" ??? l.285: are the image points in pixel units? l.310: coaxial -> binocular? l.326, 331, 394: what are a and b in the integral? l. 386: finishing -> end
Preliminary Rating. This rating indicates to the area chair, to other reviewers, and to the authors, your current opinion on the paper. Please use 'Borderline' only if the author rebuttal and/or discussion might sway you in either direction.	Weak Reject
Preliminary Evaluation. Please explain to the AC, your fellow reviewers, and the authors your current opinion on the paper. This explanation may include how you weigh the importance of the various strengths	The paper is not convincing at all: - the section on initialization is unclear - from my point of view, there are many degenerate scene configurations,

and weaknesses you described above in Q1-Q3. Please summarize the key things you would like the authors to include in their rebuttals to facilitate your decision making. There is no need to summarize the paper.	especially planes, which are not discussed - the experimental results on real scenes are very disappointing, and no comparison is done against the state of the art in multimodal matching
New exciting ideas. CVPR'16 would like to draw attention to papers that explore highly innovative ideas, novel problems, and/or paradigm shifts in conventional theory and practice. Such papers may not be "complete" in the "traditional" manner in the sense that it may not be possible to have experimental results comparing other related efforts or that they may not have large, publicly available data sets to be used for performance comparison. However, we expect these papers to be visionary by nature. Should this paper be considered under "new exciting ideas"?	No
Reproducibility. Could the work be	

reproduced from the information in the paper? Are all important algorithmic or system details discussed adequately?	There are not enough explanations on the initialization and on the actual numerical method used to solve the Euler-Lagrange equations to be able to reproduce the results
Confidence. Select: "Very Confident" to stress that you are absolutely sure about your conclusions (e.g., you are an expert who works in the paper's area), "Confident" to stress that you are mostly sure about your conclusions (e.g., you are not an expert but can distinguish good work from bad work in that area), and "Not Confident" to stress that that you feel some doubt about your conclusions. In the latter case, please provide details as confidential comments to PC/AC chairs (point 7.)	Very Confident