

IEEE Transactions on Software Engineering

Resubmission Response Letter

Paper reference TSE-2019-05-0164.R1

Associate Editor:

First, I want to apologize for the delay in this recommendation. One reviewer needed an extension to complete their review due to pandemic-related crises, and then ultimately was unable to complete it. I am basing my recommendation on the two reviews that have been submitted.

Both reviewers still want to see this manuscript eventually published. But both also mention a critical flaw of the work, omitting several highly cited foundational prior works in the review, with no clear rationale. Without including and addressing these works, they do not view the survey as complete or definitive. I tend to agree; surveys of this nature should be inclusive to the history of prior work. After all, the point of a survey is to help future researchers quickly understand the history and state of the art. That is not possible if notable, foundational, and highly cited works are omitted.

Both reviewers mention a number of other more minor revisions that are also essential.

For the reasons above, I'm going to recommend major revisions once again, and return the paper for additional review to the two reviewers and myself to verify that both experts view the survey's revised scope as adequately covering prior work.

Dear Associate Editor,

Please find attached a revised version of our submission “*Computer Vision in Software Engineering: A Survey*”. Thank you for your time and efforts in leading the review process; your comments were highly insightful and helped us improve the quality of our manuscript.

We prepared this correspondence to communicate how we incorporated each reviewer's feedback in our revised manuscript and also to address the questions and comments raised by the reviewers. We have quoted all reviewers' comments in this letter, followed by our response. In addition, we highlighted all major changes in the manuscript itself (indicated by green highlights).

Mohammad Bajammal, Andrea Stocco, Davood Mazinanian, Ali Mesbah

Response to Comments

Reviewer 1:

I would like to thank the authors for their consideration of the revisions suggested by all of the reviewers. I still believe that the work the authors are conducting is both timely and relevant to the SE research community. However, given some of the papers and issues brought up by the other reviewers, I am recommending that the authors prepare another major revision of this manuscript. The primary reason for this has to do with the current scope of the search strategy, as I explain in more detail below.

First, in general, I would like to thank the authors for addressing most of my concerns from my original review. In particular, the summary at the end of each RQ has made the paper much more readable overall. I also appreciate that the authors made the results of their work available in an online appendix attached to a GitHub repository. I would like to see this collection of papers continue to grow as this research area matures.

As for addressing some of the comments of the other reviewers, I also believe the authors did a generally commendable job, with a few exceptions as I outline below.

Thank you for the positive feedback.

I also agree with Reviewer 2 that the current title is a bit vague when taken at face value. It could be taken to mean that the paper is a survey on incorporating CV during the engineering process of a software system. One concrete suggestion for the title I might consider is “A Survey on the Use of Computer Vision Techniques to Improve Software Engineering Tasks”. I think this strikes more squarely with what the authors are trying to do in the survey.

Thank you for the feedback. We have modified the title according to your suggestion.

The authors argue that their paper is not an SLR, and generally I would agree. However, I think they need to clarify why an SLR is not appropriate in this context. Are there too few papers? Is the research area too young? Some more clarification on this point would help to solidify the overall purpose of the paper.

Despite computer vision having a long history, its application to software engineering tasks is a quite recent approach in SE. In fact, more than half of the collected papers were published within the last five years. Furthermore, with a total number of 66 papers that are spread across various SE tasks (e.g., regression testing, reverse engineering, automated UI code generation), it is not tractable or meaningful to conduct an SLR due to the recency of the field and the relatively low number of papers that are spread across many research areas. This is the reason we opted for a survey instead of an SLR.

In general, I don't have any major objections to the current research questions given the relative

recency of this field. I will say I find that RQ2 is somewhat subjective, as the authors essentially have to interpret rationale that may or may not be explicitly stated in a given paper. Given the somewhat subjective nature of this process it would be great to understand the methodology more clearly. For instance, were papers discarded from this question if no explicit rationale was given?

Thanks for the feedback. We have now clarified this process in the first paragraph of Section 4.3. For papers that explicitly stated their rationale, we recorded their rationale as-is. This resulted in the three main rationale categories we identified in the survey (namely, context-driven, ease of use, and robustness; Section 4.3). For those papers that did not explicitly mention a rationale, we analyzed the paper's full text and classified it to the closest rationale category. As you have correctly pointed out, this certainly had to involve some subjectivity, which is due to the lack of ground-truth classification of rationales, and therefore we had to come up with the above classification process.

While the construction of the query terms and Boolean combinations could have been further refined, for this survey, I believe the search query as constructed would lead to the return of most relevant papers (if the list of venues was more complete, see later comment).

I do believe this is the first major survey I have seen on the use of CV techniques to improve/automate SE tasks. I think the revised version of the manuscript does a fine job of distinguishing the novelty from other related surveys.

Thank you!

I agree with the third reviewer that it would be interesting to include a small section that surveys state of the art developer tools or approaches from industry with the primary goal of applying CV techniques to enhance some aspect of the development process. The search process for these could be completely separate from the paper search methodology.

We have added a new section (Section 4.4.4) in which we survey industrial and open-source tools and libraries that help developers build computer vision analysis or processing techniques. We describe the goal of each tool or library, and also identify which tools/libraries have been used by the papers in our pool, and which haven't been used yet. We also added a table (Table 6) summarizing these data.

My biggest current problem with the current revision of the manuscript is that the authors did not update their search methodology and inclusion criteria given some highly related papers suggested by the other reviewers. The other reviewers mentioned several papers across different venues that would be considered outside the main SE conferences and journals, yet still contained very relevant papers to the overarching goal of this survey. For instance, even in the authors own Figure 5, two of the venues with nontrivial numbers of papers (CHI and UIST) were not included in the search methodology. Given the suggestion of highly relevant papers from these venues, it is imperative that the authors revise their search strategy to include other related papers from these venues, and not simply add the papers that the reviewers suggested. This also makes the replication of the survey paper list difficult, as it is not clear why the papers from these venues were included when they were not in the list of potential publication venues given in Table 1.

Publication venues that are more focused on HCI and user interfaces have been applying vision-based techniques to SE areas/tasks likely longer than the SE field has been, given their tighter focus. As such, there is likely much that SE researchers could learn by including such papers within this survey. This would help to give a more complete picture of this cross-disciplinary research field and provide information on the successes and failures of the past work so to inform the design of future techniques that are more SE focused.

In addition to HCI-related venues, I think it would also be beneficial for the authors to consider some top ML and CV venues, such as CVPR, as there are likely to be some papers that apply CV to SE-related tasks from these fields as well. Given the interdisciplinary nature of this work, I believe this breadth is critical for collecting a representative set of papers in this area. This is my main reason for suggesting a major revision for the current vision of this manuscript. Should the authors address this concern (although this would require quite a bit of additional effort) and some other minor concerns, I would be happy to see it published.

Following this suggestion, we have now significantly updated our search methodology and added an additional stage where interdisciplinary venues from CV, ML, and HCI are included in our search. We selected the top three venues (based on the h5 index from Google Scholar) from each field, namely: NeurIPS, ICLR, and ICML for machine learning, CVPR, ECCV, and ICCV for computer vision, and CHI, CSCW, UbiComp, and UIST for human-computer interaction.

We have added a new paragraph in Section 3.4 that describes these new modifications. Table 1 now shows an updated list (in the lowest section of Table 1) that includes the new additional interdisciplinary venues. The modified search process resulted in a significant increase in the number of collected papers. More specifically, we found 17 additional papers to include, for a total of 66 papers in our survey. In the revised manuscript, all research questions, figures, tables, and analyses now refer to this final pool of 66 papers.

In the previous version of our manuscript, papers from non-SE venues (e.g., CHI, UIST) were obtained via snowballing. Given the results of the snowballing, we agree that such venues are also worth considering in the main search procedure. This is why the revised manuscript has a significantly updated search methodology where 10 venues from non-SE interdisciplinary research areas (CV, ML, and HCI) were included in our search from the very beginning. Please note that this extension required a substantial amount of time and effort on our part.

Reviewer 3:

On the positive side, R1 fixes several of R0's problems and adds new valuable information (i.e., the data the authors published on Github in Section 2.4.3). On the flip side, R1 still does not do a good job of putting the survey into the context of prior work. The most closely related work [86] is only discussed in one paragraph on the last page. This treatment of a related survey is too little too late, as both surveys overlap significantly (the previous survey focused on testing and in this "general software engineering" survey 75% of the surveyed papers are also about testing). In some sense R1 has become worse than R0 because R0 still had a measure of this overlap (11 of the discussed papers) but R1 conveniently omits this information.

To me, this article should start by summarizing the findings of the earlier survey and then formulate its research questions in relation to the earlier work (which questions did [86] not answer and which questions need additional evidence?).

Related to the previous point, the first contribution bullet in section 1 ("first survey on the use of computer vision (CV) for software engineering") is misleading, as testing is clearly part of software engineering and we already have the earlier survey [86].

As per your feedback, Section 2 now clarifies at greater length, and from the very beginning of the manuscript, how our work has a different scope compared to Saraben et al. [86] and other works. Section 2 discusses this point in more detail, but we summarize the main differences as follows.

The work by Saraben et al. answers the following question: what approaches have been used to conduct cross-browser regression testing. In contrast, our work is not concerned at all with that problem. Our work answers the following question: in what ways have computer vision techniques been used to advance software engineering tasks. The only way in which both works are remotely related is that some of the papers found by our survey also happen to be found by Saraben et al. as well.

Other than that, both works are unrelated and do not overlap in neither their scope nor their objective. Saraben et al. focus on a specific problem (i.e., cross-browser regression testing), regardless of what approaches were used (e.g., DOM analysis, state space navigation, visual analysis). That is, the survey in Saraben et al. is problem-specific but approach-agnostic. In contrast, our survey is approach-specific but problem-agnostic. We focus on a specific approach (i.e., computer vision techniques), but consider its potential for any area of software engineering (e.g., testing, maintenance, development, design, requirements).

In addition, as per your suggestion, we have now restructured the manuscript in order to begin with a significantly expanded Prior Work section (Section 2). Previously, we had a smaller Related Work section at the end of the paper. In the revised manuscript, the section has been substantially expanded: it now describes that state of the art in three subsections concerning (1) visual GUI testing, (2) computer vision-based surveys, and (3) interdisciplinary secondary studies in SE. We hope that our revised version of Section 2 now better places our study in the context of the prior work.

While R1 improves Section 2 by providing some definitions, it still misses a definition of the “construction” box (right top of Figure 1). For example, does taking a screenshot of a UML diagram count as “construction”?

Related to the previous point, why does Section 2.2 explicitly exclude UML modeling documents, but Figure 1 shows them? The text includes taking a screenshot from the user interface, but taking a similar screenshot of the UML diagrams are out of scope? The bottom line is that the methodology is still confusing and needs clarification.

Part of the confusion is writing such as: “Further details on what is or is not included in this survey are covered in sections 2.4.1 and 2.4.1.”

While Section 2 tries to exclude some earlier work from the scope of the paper, RQ1 is written very broadly to include all aspects of software engineering. Maybe RQ1 should be reworded to the scope the article intends to have?

The question of in what ways can visual data be constructed is an important one. This is why our RQ3 (Section 4.4) focuses entirely on this question, and explores in detail exactly how computer vision has been applied to construct and process visual data from software.

We agree that the statement regarding UML could have been written in a better way. It was included from an earlier revision to clarify, for another reviewer, that our work is not concerned with techniques of how to *visualize* UML documents. We have now rewritten the Scope section (Section 3.2) in a clearer way.

As for RQ1, we do not exclude any SE areas or tasks from our survey. As can be seen in Figures 4 and 6, the scope of our survey includes various areas of SE (e.g., testing, requirements, development, design, maintenance).

While R1 is more comprehensive than R0, R1 still omits important early papers (despite the letter to the editor claiming otherwise). Omitting such early foundational work is just not right. They have been highly influential and attracted hundreds of citations.

After feedback from reviewers, we have now significantly updated our search methodology and added an additional stage where interdisciplinary venues from CV, ML, and HCI are included in our search. We selected the top three venues (based on the h5 index from Google Scholar) from each field, namely: NeurIPS, ICLR, and ICML for machine learning, CVPR, ECCV, and ICCV for computer vision, and CHI, CSCW, UbiComp, and UIST for human-computer interaction. We have added a new paragraph in Section 3.4 that describes these new modifications. Table 1 now shows an updated list that includes the new additional interdisciplinary venues. The modified search process resulted in a significant increase in the number of collected papers. More specifically, 17 additional papers were found, for a total of 66 papers.

Having said that, we already included the work of Dixon et. al. 2011 (paper #52 in the previous version) in our existing manuscript. The paper you mentioned by the same authors is a minor

delta (from the perspective of our survey RQs) that did not add new findings. However, for the sake of completeness, we have added the papers you mentioned to the pool. This is, of course, in addition to papers from the updated search methodology explained above.

(Minor issues)

"SRL" should still be "SLR". Also the article should still define "SLR" before using it.

Figure 4 line types are still very hard to distinguish, which makes the figure still very hard to understand.

Thank you for pointing these out. We did another pass on the paper to fix typos. As per your suggestion, we also improved Figure 4 and added various types of lines and marker shapes to make it easier to distinguish the different trend lines.