

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Daniil Kononenko

PhD Program: Computational and Data Science and Engineering

Title of Thesis: Gaze redirection in Images Using Maching Learning.

Supervisor: Professor Victor Lempitsky

Chair of PhD defense Jury: Professor Maxim Fedorov

Email: m.fedorov@skoltech.ru

Date of Thesis Defense: October , 2017

Name of Reviewer:

<p>I confirm the absence of any conflict of interest</p> <p>(Alternatively, Reviewer can formulate a possible conflict)</p>	<p>Signature:</p>  <p>Date: 17-10-2017</p>
---	---

The purpose of this report is to obtain an independent review from the members of PhD defense Jury before the thesis defense. The members of PhD defense Jury are asked to forward a completed copy of this report to the Chair of the Jury at least 30 days prior the thesis defense. The Reviewers are asked to bring a copy of the completed report to the thesis defense and to discuss the contents of each report with each other before the thesis defense.

If the reviewers have any queries about the thesis which they wish to raise in advance, please contact the Chair of the Jury.

Reviewer's Report

Review of the PhD dissertation of Daniil Kononenko's "Learnable warpin-based approach to image re-synthesis with application to gaze redirection".

Background and objectives

The main objective of this Doctoral Thesis is to propose new machine learning based approaches for image re-synthesis problem. In particular, the work concentrates on gaze redirection problem, which aims to re-synthesis a novel photorealistic output from the given input image in such a way that the gaze of the target person in redirected by a certain angle. This problem is relevant in many practical applications, such as video conferencing where eye-to-eye contact is often lost due to vertical gap between the display and the video camera. Other relevant applications include photo and video editing, where gaze might need to be redirected to be consistent with the ideas of the photographer or movie director. Many

companies are working on the solution to this problem, which further emphasises the general interest towards gaze redirection. However, even though gaze redirection problem has been studied for several years, there exists no methods that produces satisfactory results from monocular input images and is applicable in real-time scenarios.

Kononenko's thesis concentrates on re-synthesis problems where the desired transformation is defined by a dataset of pairs of input and desired output images. The proposed approaches use warping field concept, where each output image pixel is sampled from the input image. The predictor for the warping field is learned from the given dataset. In gaze redirection problem, the warping field is constructed only for the eye region in such a way that the gaze direction is altered by a given angle in the output. This is slightly different from the mainstream approach where the output is generated by synthesising entire image from a novel camera viewpoint. The selected approach avoids several problems related to e.g. dis-occlusions. Moreover, the selected approach allows efficient implementations suitable for real-time applications.

The proposed approaches include two random forest based methods and two neural network approaches for learning the warping fields. The candidate has carried out a comprehensive literature review, which covers the most recent and important re-synthesis approaches. The proposed solutions are technically sound and they have also appeared in one journal (TPAMI) and two conference articles (CVPR and ECCV). The candidate is the main contributor and the first author in two of these works (TPAMI and CVPR). He has also substantial contributions on the third article as described in the manuscript. In addition, the work has result in a patent and the technology has been licensed to a company. These further emphasise the practical value of the thesis.

Structure of the thesis and the contributions

Chapter 1 presents image re-synthesis and gaze redirection problems with related literature review. In addition, the final part of the chapter lists the contributions and anticipated impact of the thesis work. The literature review is comprehensive and covers the most important related works with reasonable details. The structure of the chapter is slightly complicated and could benefit by splitting the discussion of related work into further subsections or paragraphs. A table summarising the previous work on gaze redirection with their main properties could be useful as well. The caption of Figure 1.14 was slightly confusion and could be rephrased. The contribution of the thesis and the author's role in the related publications is outlined clearly.

Chapter 2 begins by formulating the re-synthesis problem as a pixel-wise replacement and describing the general pipeline for image re-synthesis. The following sections introduce publicly available Columbia Gaze dataset and a new Skoltech dataset. Skoltech dataset contains videos of 150 individuals with known gaze directions. The collection procedure of the dataset is well described. It would have been interesting to have additional discussion about effect of camera position and a possibility to utilise eye tracker in data collection. All videos seem to be recorded with same distance from the camera. The effect of distance could have been discussed and perhaps evaluated further. It was also unclear if the dataset was collected as a part of this thesis work. In such case, it could have been clearly listed as a contribution in Chapter 1. Finally, the last part of Chapter 2 discusses face alignment approaches and general approach for gaze redirection.

Chapter 3 presents a weakly supervised random forest approach for obtaining warping fields for image re-synthesis. The chapter begins with general introduction to random forests and their applications in computer vision. The latter part of the section presents so called warping flow forest method, which is one of the main contributions of this thesis. Warping flow forest

learns to predict a warping field for a fixed gaze redirection angle. The chapter contains good description of the random forest architecture and the training procedure. The method runs in real-time with standard CPU and produces reasonable results. To my knowledge, similar warping field based solution to gaze redirection has not been presented before. The drawback of this approach is that it learns to redirect gaze with fixed angle and it has relatively large memory footprint as the author also points out. In addition, it would have been interesting to have further discussion about the temporal stability of the method, how face registration affects the practical performance of the method, and a short description how the numerical parameters were obtained. Finally, Chapter 3 presents experimental evaluation of the proposed method. Evaluation is well written and the results are clearly presented. The main challenge is to understand how the differences in the MSE errors reflect the practical performance differences of the methods. In my opinion, the work could have included slightly more comprehensive discussion about practical validity of the quantitative evaluation criteria.

Chapter 4 presents a deep learning based approach called DeepWarp for image re-synthesis. Similarly, to the warping flow forest, this approach also learns to predict the warping field that is applied to construct the final output image. As the author points out, his contributions to this approach is limited to experimental setup, data preparation, and the implementation of the user study. DeepWarp approach is described well, although the description of the interpolation kernel and related backpropagation on page 63 could be slightly elaborated. The experiment part of the Chapter is comprehensive containing both quantitative and user study evaluation of the approach. Description of the models in Section 4.6.1.1 could have been more comprehensive to help the reader to understand the modifications done to DeepWarp architecture (particularly 2-4). The setup in the user study is very interesting and provide clear additional value of the MSE based approach. Perhaps some discussion about the limitation of the user study approach could have been included. For instance, it only evaluates only the photorealism of the results and as far as I understand a trivial method that produces zero warping field would get high score. Finally, there could have been discussion about the possibility to include similar lighting correction module also to random forest approach.

Chapter 5 introduces another random forest based approach for predicting the warping field. In this case, the forest is trained using teacher-student setup where the DeepWarp model (Chapter 4) is utilised to train the random forest. The upside of this approach is that the forest can be trained using the actual warping flow, which is not directly available in the original dataset. This setup leads to a random forest that has better performance and clearly smaller memory footprint compared to the approach presented in Chapter 3. The idea is very good and it seems to result in a method that is both efficient to compute in practice and produces high quality results. Only drawback is that this method also works with fixed redirection angle. Nice illustrations of the resulting warping field are presented in Figure 5.1., similar could have been produced also in previous chapters. The discussion about the actual gaze redirection at the end of Section 5.3.1. was interesting. Figure 5.3. could also include the results from the constant warping field baseline.

Chapter 6 presents semi supervised approaches to gaze redirection problem. Chapter begins with discussion about the related work utilising image analogies. The discussion is easy to read and is comprehensive enough. Chapter continues by introducing the proposed learning architecture, which is very interesting and novel at least in gaze redirection domain. The main advantage of the approach is the ability to utilise unlabelled video sequences in training. The latter part of the chapter contains experimental evaluation, which indicates that unsupervised approach results in lower MSE error if number of training data is limited. However, it was not easy to interpret the actual difference in the practical quality from the small differences in MSE errors. In addition, it would have been interesting to know if gaze direction estimation network could have been utilised to evaluate the analogy property (second paragraph in Section 6.4).

Chapter 7 contains overall discussion and conclusions of the thesis. The summary is clear and the discussion is relevant and easy to read. The limitations regarding the gaze redirection by re-synthesising only the eye vicinity is interesting (first paragraph in Section 7.1.) In my opinion, these could have appeared already in earlier chapters.

Presentation of the results

The problem settings, previous works and the state of the art in the research field are presented in the thesis manuscript in sufficient length and detail so that the interested reader of the dissertation can get proper understanding how the new contributions advance the state of scientific knowledge. The most important findings are discussed in the text, and conclusions based on these findings are sound. The presentation of the background and the results thus follows good scientific practices. The use of literature references and their appearance at the end of the manuscript mostly follows the scientific practice. The English language of the manuscript is mostly very good. The text is pleasant to read and the evolution of the presented ideas is for the most part easy to be followed.

Conclusions

In summary, the overall impression is positive as the problems addressed are topical and the proposed approaches provide new practical solutions to well-known computer vision problem. Based on the above examination of the manuscript of Daniil Kononenko, I recommend that the candidate should defend the thesis by means of a formal thesis defence.

In Rovaniemi on 16th October 2017



Esa Rahtu, Ph.D. (Tech.)
Assistant Professor
Department of Signal Processing
Tampere University of Technology
Finland

Provisional Recommendation

I recommend that the candidate should defend the thesis by means of a formal thesis defense

I recommend that the candidate should defend the thesis by means of a formal thesis defense only after appropriate changes would be introduced in candidate's thesis according to the recommendations of the present report

The thesis is not acceptable and I recommend that the candidate be exempt from the formal thesis defense