Cover Letter

Isidoros Marougkas

Profile:

I am a soon-to-be Electrical Engineering graduate of National Technical University of Athens, Greece, with specialization in AI, Control Systems and Robotics, with high grades, a proven track of undergraduate research experience, as well as self-organization and leadership abilities. Just on the verge of completing this chapter of my studies, I am actively searching for a fixed-term research residency position related to either of those areas.Being aware of the level of international competition at my field, my goal is to join a professional destination that will enhance my practical research experience, teach me how to cooperate in an interdisciplinary and multicultural environment and strengthen my publication resume, as I, ultimately aim to apply for a top-notch PhD program in the future or become member of a worldwidely impactful research team.

A brief explanation of why I consider myself an ideal candidate for your position:

- excellent theoritical Mathematical, Physics and Algorithmic background
- acquainted research experience in handling a Computer Vision problem (6D Object Pose Tracking) using Deep Learning techniques for the past 18 months with measurable results (76% reduction of the mean State of the Art error metrics)
- practical implementation of Computer Vision, Sound Processing, Deep Learning, Robotics and Control projects in Matlab, Python, PyTorch
- upcoming submission of a manuscript with the intend of publication in the ICIP 2020 conference along with open source code release
- consistently outstanding academic performance and leading ability throughout the years
- active member of student organizations and forums, proven leadership abilities, trained soft skills
- soon to be Electrical Engineering graduate of the top-ranked school in my country (Feb. 2020), open to fresh project ideas and willing to relocate (25 years old)

And now, let me talk about my career trajectory and potential aspirations, in a little more detail:

At the moment, I am an undergraduate Electrical and Computer Engineering student at the National Technical University of Athens, having successfully completed all 60 courses required on time and conducting research in the intersection of Computer Vision and Machine Learning, occasioned by my Master Thesis for the last 18 months, aiming to graduate by the end of February, 2020 in order to obtain my Diploma (5-year joint B.Sc. & M.Sc., 300 ECTS) with a GPA of 8.88/10 (around top 6% between classmates). I was admitted 48th among 350 students in the aforementioned school, the highest ranked Engineering school in Greece, after achieving an excellent performance at the National University Admissions Exams, scoring a total of 19.054/20.000 (top 0.5% among 80.000 students nationwide), for which I was nationally awarded. Since the very first years of my undergraduate studies, I have been an active member of the student society, helping in the administration of the students' forum and been selected for funded Educational Trips and Summer Schools. At the same time, I have demonstrated leadership qualities in coordinating a volunteering team that helped organizing the biggest career fair in Greece, - in continuation to my high school years, when I co-founded the institution's theatrical group, that is still active today. As for my academic performance, my proclivity for advanced mathematics and physics allowed me to achieve excellent marks in almost all the relevant courses (Calculus, Linear Algebra, Differential Equations etc.) providing me with solid fundamentals that have greatly facilitated my further studies. I have always had the desire to complement abstract theoretical concepts with equivalent real-world applications, so when the time to select my Major came, I deliberately gravitated towards the direction of Computer Science and specifically a compilation of Software and Hardware systems, Signals, Control Theory, Robotics and Machine Learning relevant courses. Specifically, I focused my attention towards the latter four and I chose to delve as deeply as my University curriculum allowed me to concepts such as Artificial Intelligence, Pattern Recognition, (Deep) Neural Networks, Computer Vision, Digital Signal Processing, Linear/Non-Linear/Optimal Control Systems Analysis and Design and Robotics, ranging from their theoretical foundations to hands-on implementation projects.

When the time for conducting my Master Thesis came, I did not just want to get on with it as soon as possible, but, on the contrary, I saw it more as a challenge for myself to prove that I can rightfully belong in the AI research community and that I can efficiently combine previously established knowledge to produce publishable results. This led me to join the CVSP group of NTUA and conduct my Master Thesis on Temporal 6-DOF Object Pose Tracking with the use of Deep Learning techniques under the supervision of Prof. Petros Maragos. Visual Pose Tracking finds a variety of further research and commercial applications in Robotics, Activity Recognition and, most usually, Augmented Reality. However,there are plenty of challenges that come with trying to solve this problem: an efficient 6-DOF Tracker must be robust to background clutter, occlusions, sensor noise, appearance and illumination change. It must, also take into consideration pose ambiguities that have a dual source: the representation choice of the 3D rotation space and the symmetries that arise from the object's 3D geometric model. Last but not least, Pose Trackers must handle small drifts that accumulate over time, as well as, think about long-term trajectory dependencies before providing a pose prediction.

On top of all those, employing a tracking algorithm as a component of a broader application requires real-time speed and limiting the dependencies on the depth modality as much as possible, as it is not available to all devices. Most frequently in literature, this problem is formulated as a learning one as follows: a synthetic dataset of RGB-D rendered pairs of the object set at two distinct, but close, poses is built. Those pairs are used as training samples in a two-stream Convolutional architecture that, later, are merged in order to output their pose difference and are trained by minimizing a standard regression loss in the Euclidean space[2]. During inference, the one stream inputs the actual RGB-D video clip that comes out of the sensor and the other one processes the corresponding pair that is created by closing a feedback loop rendering the predicted object pose at the previous timestep. Most such approaches either neglect some of the aforementioned challenges (e.g. using Euclidean losses to minimize non-Euclidean rotational distances), hope that the model will make sense through an indirect data augmentation scheme or incorporate the pose matching architecture, explained above, on top of slow architectures (e.g.using deep MaskR-CNN[3] base for foreground/unoccluded pixel segmentation to tackle clutter/occlusions,correspondingly).

During my research, I implemented important alterations to this Deep Learning strategy that reduced both the mean State-of-the-Art translational and rotational errors by an average of 76% when the tracker is tested in the hardest dataset available in the field: an object moving freely in 3D space and dynamically interacting with the user's hands, without violating real-time constraints. In detail, my main contribution was to handle background clutter and occlusions via two equivalent parallel Soft Spatial Attention modules that enhance the pixels of the video stream features that belong in the object and are not occluded by the user's hands, before the pose comparison intended stream merging. Those modules were trained by minimizing two Binary Cross Entropy auxiliary losses, that used binary masks as supervising signals. These Attention modules, also, provide intuitive explanations of the conditions under which the tracker loses the object and of which object parts it focuses the most on when it does not. All the Multitask sublosses were weighted by parameters optimized during the Network's backward pass. Furthermore, instead of the classical Euclidean regression loss functions, I used a 3D Geodesic Rotation loss that respects the geometry of the Riemannian SO(3) rotation space, employed a continuous rotational parametric representation that facilitates network training and utilized the object's 3D geometry to assign different weights to each rotational component.

Currently, I am cooperating with doctoral researchers from the CVSP group, writing up a manuscript of this work that will be submitted to the International Conference of Image Processing (ICIP) 2020. At the same time, I am experimenting with some possible extensions of my idea in order to take care of the rest of the tracking problem's challenges, aiming to further publications. At first, for long-term dependency handling, I am trying to model the time-component of the problem by replacing distinct pose differences with randomly sampled 3D continuous trajectories where the translational components are derived by cubic splines and the rotational ones by geodesic paths in SO(3) and process them using Convolutional-Recurrent units. Intending to counter short term drift, I am experimenting with a hierarchically Attentional Optical Flow distillation scheme that will guide the latent features towards higher tracking precision. Finally, aiming to cut out the need for Depth maps during inference, I am trying to add to the overall architecture Pixel-Adaptive Convolutional modules that distill depth information from computationally free Depth samples, available only in training, and bring solely RGB-based pose predictors on par with RGB-D-based ones.

This prior experience has provided strong proof of my skills even to myself and has ignited the spark for AI research to become the center of my attention. It has grown both my confidence and curiosity to discover and solve novel open challenges. It has taught me how to take an approach that already is at a certain stage and adapt working achievements in other areas by combining them with intuitive insights of the problem to, ultimately, overcome the given performance level. But most importantly, it has shown me that research is more about small, painstakingly detailed steps and less about humoungus epiphanies. Knowing that has facilitated me to properly direct my work habits, as well as, to handle the psychological ups and downs that are caused by inevitable intermediate failures.

My general research motivation is to bridge sectors of Artificial Intelligence, namely Computer Vision, with Control architectures of Robotic systems. The most intriguing example of doing so that comes to my mind is Automated Vehicles. That has a two-fold explanation: One shallow and technical, and one deeper and ideological. The former relies on the fact that Autonomous Driving is the major concept of AI, at the moment, that combines the majority of Computer Vision subfields, ranging from Human Pose Estimation to Traffic Sign Recognition and Adversarial Attacks. Yet, these attempts do not stick to mobile cameras, but also explore fusion strategies with other kinds of sensors that provide different kind of information or that fail under different conditions. The way this problem will ultimately be handled (e.g. if more weight will be put in the accurate Perception or the robust Planning phase of the navigation system, if those systems will rely more on expensive and accurate low-resolution LiDAR or cheap and high-resolution cameras and if the use of 2D or 3D convolutional architectural concepts will, finally, be proven to be the most efficient etc.) will pave the way for addressing other open AI problems. The latter is based on my desire to contribute to the demise of one of the most prominent causes of misery in the world. Today, every 23 seconds a death occurs on the road somewhere in the world, while my country, Greece, recently was announced to be the member of the EU with the highest risk of car accidents. I feel genuine anger towards those statistics and I strongly desire to put my mind and efforts into diminishing them. Besides, fully (Level 5) or -at least- highly (Level 3 or more) automated vehicles will help optimizing traffic, a lighter matter, of course -but significant one, no doubt- as it consists both an economic burden and a source of degradation of the quality of life in large cities. Lastly, I want to actively take part in the ethical and philosophical controversies that will arise during the design of such autonomous systems (e.g. the ‘’Trolley problem”) and conduct research that helps to co-form the conclusions of such complex socio-technical debates.

Furthermore, I am extremely curious about the applications of Machine Learning in extracting contextual information from films, as they cover the whole audiovisual spectrum. More specifically, I have been studying about the recent advances in Movie Summarization and Unsupervised Sound/Image information distillation from a Video/Sound stream, correspondingly, and I deeply wish to engage in a relevant project in the future. That is because I consider that the extraction of high level pieces of information, such as sentiments, without deliberate guidance, consists the Holy Grail of bridging the gap between intelligent systems and humans. I, also, would enjoy delving into Natural Language Processing projects, especially those related to Visual Question Answering or Political Text Analysis (e.g. fake news detectors, a topic that became fashionable in the end of the previous decade). Last but not least, I always fancy reading about new developments in Reinforcement Learning applications in Robotics, such as the recent achievement of OpenAI's team to teach a robot hand to dexterously manipulate a Rubic's cube[1], that recently took over the Internet by storm. I reckon such projects increasingly fascinating on the one hand, as they imitate human brain behaviour more closely than their supervised counterparts and, on the other, participating on a project like would utilize my previous academic background to the maximum.

Nevertheless, the above examples are far from binding. I have both the desire and the background necessary to get involved with any interesting research project idea in the fields of AI, Robotics and/or Autonomous Control that may be presented to me.

Consequentially,accepting me as a Facebook resident would mean that I would be given the chance to work in some of the best infrastructures in the world, use the most advanced computational tools and be part of an organization that is both ambitious in achieving long-term scientific goals, as well as making an impact in today's world. I hope that inside the program I will find the proper mentorship that will allow me to cultivate my skills even further, based on my pre-existing experiences and work ethic, and, ultimately, reach my potential to fulfill my scientific curiosity. But, above all, I value the opportunity of working in an interdisciplinary environment full of some of the brightest minds of my generation. I expect not to be the de facto smartest man in any room during my residency and I look forward to challenging my values and methodological habits out of my comfort zone. On the other hand, your team would integrate a proven young researcher with well-balanced analytical and coding skills who is passionate about extending the boundaries of research.

References

[1] Ilge Akkaya et al. "Solving Rubik's Cube with a Robot Hand". In: *arXiv preprint arXiv:1910.07113* (2019).

[2] Mathieu Garon and Jean-François Lalonde. "Deep 6-DOF tracking". In: *IEEE transactions on visualization and computer graphics* 23.11 (2017), pp. 2410–2418.

[3] Kaiming He et al. "Mask r-cnn". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2961–2969.