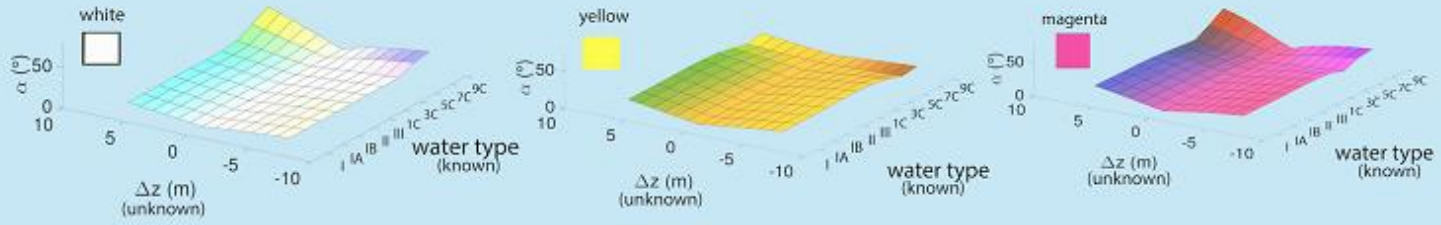
CVPR DAILY
July 21-26 HONOLULU 2017
Computer Vision & Pattern Recognition

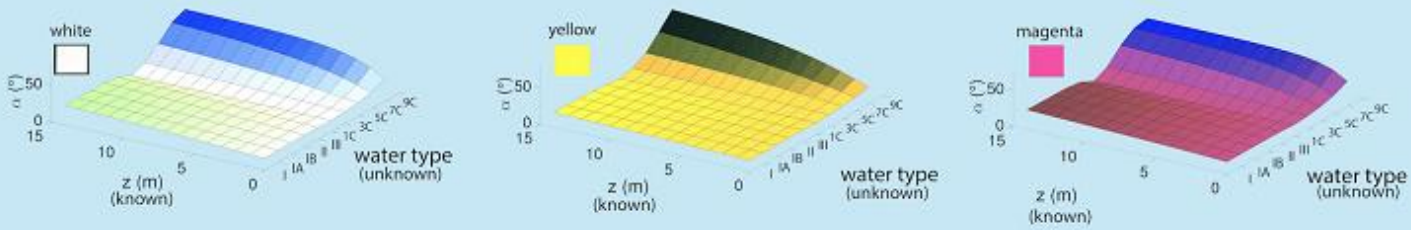
Saturday 22

ERRORS FROM INCORRECT COEFFICIENTS

Calculating β_c from incorrect range...



... or incorrect water type causes strong hue shifts in color reconstruction.



A Word of Welcome from:
Anthony Hoogs
General Chair, CVPR 2017

EXPO:
Modar Alaoui
CEO of Eyeris

Presenting work by:
Derya Akkaynak
Silvia Vinyes
Roey Mechrez
Dan Xu

Women in Computer Vision:
Ilke Demir, Facebook

Tobi's Picks and **More...**

In cooperation with
Computer Vision News
The Magazine of The Algorithm community

A publication by
RSIP VISION

Tobi's Picks

CVPR DAILY July 21-26 2017 HONOLULU Saturday

For today, Saturday 22

Tobias Baumgartner - Yahoo CV&ML team

"I received my M.Sc. in CS with a focus on ML and CV from RWTH Aachen and started my career at a small Computer Vision startup in Berkeley, CA: IQ Engines. Shortly after joining, we got acquired by Yahoo, where I now work as a Sr. Research Engineer. We build the image intelligence behind various Y! properties like Flickr, Tumblr or Y!Mail, as well as doing independent research. Currently I am mainly focused on face detection, recognition, clustering, and attribute extraction and am looking forward to gaining new inspiration for my research at CVPR'17."

These are Tobi's picks for today, Saturday 22. Don't miss them!

Morning:

S1-1A.8 09:28 Page 5 of the Pocket Guide:
On Compressing Deep Models by Low Rank and Sparse Decomposition

P1-1.46 10:30 Page 7 of the Pocket Guide:
Detecting Masked Faces in the Wild With LLE-CNNs

P1-1.73 10:30 Page 8 of the Pocket Guide:
Subspace Clustering via Variance Regularized Ridge Regression

Afternoon:

S1-2B.13 13:30 Page 12 of the Pocket Guide:
Crossing Nets: Combining GANs and VAEs With a Shared Latent Space for Hand Pose Estimation

S1-2C.30 13:30 Page 13 of the Pocket Guide:
TGIF-QA: Toward Spatio-Temporal Reasoning in Visual Question Answering

O1-2B.22 14:17 Page 13 of the Pocket Guide:
Disentangled Representation Learning GAN for Pose-Invariant Face Recognition

P1-2.83 15:00 Page 15 of the Pocket Guide:
Stacked Generative Adversarial Networks

"Besides papers, while on Oahu I want to try and soak in as much nature as possible. I will be exploring the coast line by bicycle (I am especially curious to ride around the north shore) and try to run as many hikes as the evening hours allow. Quick head's up - Diamond Head closes at 6p. Overall I want to get in 50mi of running, as I prepare for my first ultra marathon."

Aloha, CVPR!

Welcome to **CVPR 2017** and welcome back to of **CVPR Daily**, the magazine launched one year ago by **CVPR** and **RSIP Vision** at CVPR 2016 in Las Vegas.

Also in this CVPR 2017 edition, you will read great stories, reports and articles about the **impressive scientific work** being presented here in **Honolulu, Hawaii**. Today and for 3 more days we will do our best to present our readers a little bit of the **immense talent** on display at the most important computer event of the year. Please do not miss the word of welcome of the General Chair at page 6.

On behalf of CVPR, RSIP Vision and Computer Vision News, **I wish you an awesome CVPR 2017!**

Ralph Anzarouth
Editor , **Computer Vision News**
Marketing Manager, **RSIP Vision**

CVPR Daily
Publisher: **RSIP Vision**
Copyright: **RSIP Vision**

Our editorial choices are fully independent from CVPR.

Derya Akkaynak

CVPR DAILY
July 21-26 HONOLULU 2017 Saturday

What Is the Space of Attenuation Coefficients in Underwater Computer Vision?

Derya Akkaynak is a postdoctoral researcher at the **University of Haifa**, jointly with the **Interuniversity Institute for Marine Sciences in Eilat**.

"The other deep in computer vision"

Derya (whose name means '*ocean*' in both Persian and Turkish) told us about her work which she is presenting today, titled "**What Is the Space of Attenuation Coefficients in Underwater Computer Vision?**".

The poster she is presenting is about improving color reconstruction and color acquisition in images that are collected underwater with underwater robots or divers. Derya, who is an oceanographer and mechanical engineer (not a computer scientist by training), is working on understanding how light propagates underwater and how it gets captured on camera sensors, to find out how we can compensate for the colors that are lost, in an accurate and objective way. Her paper leverages decades worth of data from optical oceanography to improve underwater computer vision algorithms, bridging the two fields.

A main challenge of her work is to validate the mathematical models they build, and see if they actually work in an underwater setting. This requires several dives, a lot of equipment, and a lot of hardware and sometimes things can go wrong, or the results are unexpected, and then they have to go back to the model and adjust it. "*So it's a constant iteration between work on the computer, and work in the sea*", Derya says, which is different to most Computer Vision fields, where work is mostly done on a computer.

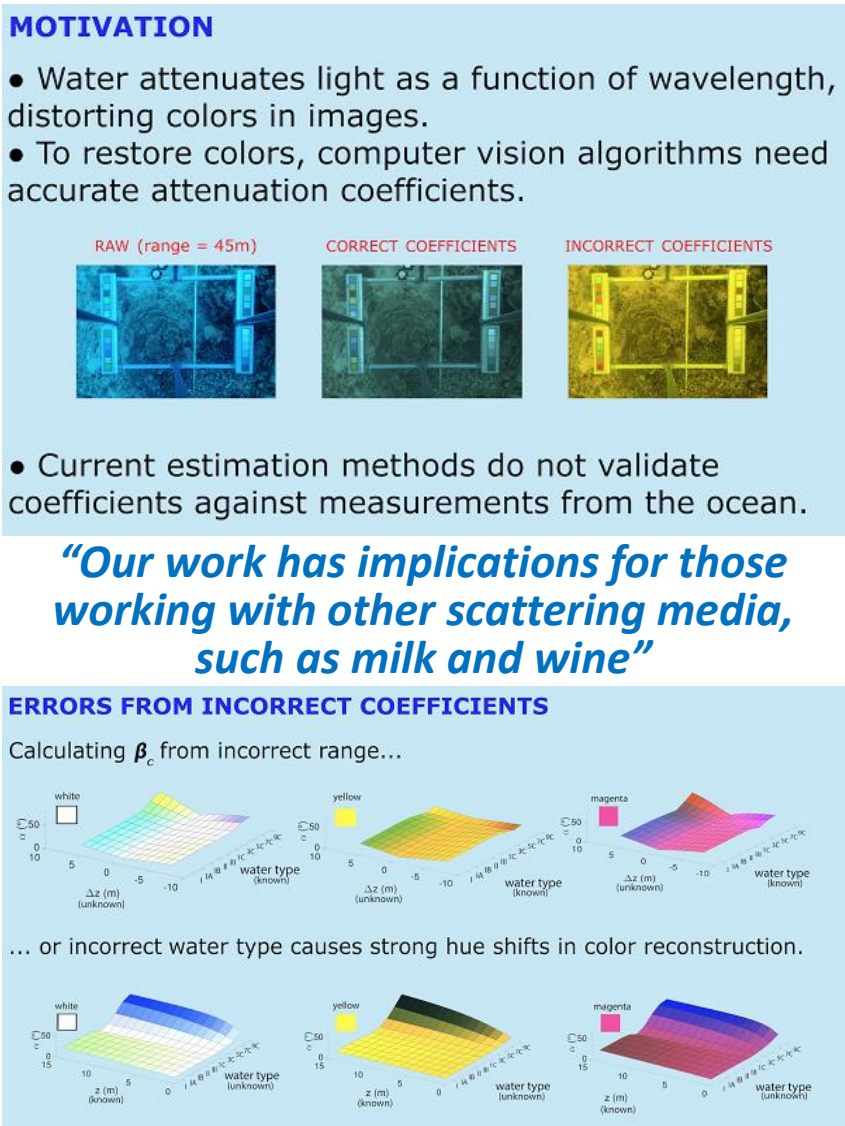
Talking about previous work, Derya explains to us that there is an existing system of equations for underwater image formation, and that everybody uses these equations. However, Derya and her co-authors looked at how these equations were derived and they found that due to two simplifying assumptions there are errors that are introduced that affect people's work in color reconstruction. So instead of using what was commonly accepted, Derya and her co-authors questioned it and were then able to highlight the weaknesses, and offer a better solution.

We asked Derya if she thinks that we can one day see underwater images just like we see things in normal life. She hypothesizes that this might be possible with a lot of specialised equipment, because we now understand very well what happens to light and how cameras capture light underwater.

So it's possible to un-do that effect, but she is unsure if it will be commonplace enough to just put on a mask and that this mask will compensate for everything. She thinks that such a mask will probably complicate diving a bit, and might be more interesting for commercial applications.

"*Our next step is to now derive a new system of equations for underwater image formation*", Derya says, "*to compensate for the two weaknesses that we found*". They want to be able to tell people what kind of errors they should expect when they use the old equations, and they can then judge how good their results are when all the errors are taken out.

For their work, Derya and her co-authors used image formation equations that are known to be used for camera and image simulation.



"Our work has implications for those working with other scattering media, such as milk and wine"

What was important and key in their work, and their biggest contribution, is that they have brought over eight decades of knowledge from optical oceanology into Computer Vision. Because in Computer Vision, researchers estimated some coefficients that they use to correct colors. "*But they never checked if those coefficients actually make sense given ocean conditions*", Derya says. And in Oceanography it is known that light attenuation changes by place and by time, which oceanographers have mapped for the last eight decades with various instruments, from very simple to very technical. So what they have done in this work together with Tali Treibitz is to bridge Computer Vision and Oceanography.

Derya also told us about what she thinks is the biggest misconception about underwater imaging: "*In almost 99% of underwater Computer Vision papers I read, people state that light attenuated unevenly underwater; red attenuates faster than blue or green*". And this is exactly what happens in waters that are dominated by plankton (small plants that drift in the water), she explains. But if you go to coastal water, where there is contamination from rivers, soil, sediment and other non-organic substances, actually blue attenuates faster than red.

Do you want to learn more about Derya's work? Visit her poster today (Saturday) after 10:30 - Poster Session P1-1.

CVPR DAILY July 21-26 2017 Honolulu Saturday

A Word of Welcome from the General Chair

"The survey is data, and data speaks very loudly"

Anthony Hoogs is the Senior Director of Computer Vision at Kitware, a small software R&D company that does a lot of computer vision. For the CVPR conference, he is one of three general chairs, the one who is in charge of the convention center, the EXPO logistics site here in Hawaii.

It must have been very exciting to prepare for this day.

Very exciting! We've been planning this for 3 years. We did not plan to be in Hawaii. We planned to be in Puerto Rico. Switching a year ago was somewhat expensive and not really disruptive, but it wasn't that bad because most of the logistical preparation and the detail level hadn't really yet begun. Switching to here primarily meant rearranging rooms, coming out here for a site visit, and so on…

Primarily we've been increasing the professionalism of the conference through support contractors. This year we brought on an EXPO support contractor called Hall Erickson who has been doing very good things for us to manage the EXPO. We have C to C Events, who's been growing with us and manages registration, logistics of the conference, hotel contracts, and all of that.

As a general chair, my main job is to make sure that overall things are getting done, make sure that we didn't forget anything, and track the important things, but mostly there's a groups of probably 25 or 30 people now who really run the conference across all the range of things that we do. I need to make sure that those people get assigned and know their jobs.

I understand that this is going to be the biggest CVPR conference ever. Do you have any numbers that we can share with our readers?

As of this moment, we have 4,880 registrations. Last year, there were 3,650 I believe. The increase is approximately a little more than 33%. We usually get a bunch of walk-up registrations as well so we may get over 5,000. This increase has actually been planned for. If you look at the year-over-year growth of CVPR over the last few years, it has been typically

about 15% each year. We expected a bit of a bump coming to Hawaii, and the growth rate is probably more than linear. We planned ahead and reserved enough space for up to even 6,000 people in the entire convention center. We could have handled that. Handling 5000 should not be a problem, and we have enough space for the posters and the growth of the EXPO.

The EXPO has grown by almost 30% this year in the number of companies. It is almost 80% bigger in terms of sponsorship dollars. Everything is bigger. We have had more than 30% more submissions and about 30% more papers in the program so it means more poster space. There's more workshops. We received 63 workshop proposals and accepted, I think, 44. I think there is 21 tutorials. Everything is bigger, but we've been able to accommodate it all so it shouldn't be crowded.

What are the novelties of this year?

This year we have a few primary new elements in the conference that perhaps have been overdue and some of them have been talked about. All of them were discussed repeatedly in our original proposal 3 years ago to the community and have been reinforced since. First and foremost, we went with 3 parallel tracks for oral sessions. The conference has had 2 parallel tracks since 1991 when there were about 270 people. We have not grown the number of tracks as other conferences have done over the years. Some conferences have not, like ICCV, but we felt it was time to expand the number of tracks and have correspondingly more oral presentations. We are doing that for 2 days actually. The final day-and-a-half of the conference is just 2 tracks as it has been.

The other new element is the three-and-a-half day conference. The main conference lasts 4 days, as it has been done recently and has been very popular in our surveys. However, we now have a half day on the 3rd day which gives people a bit of a break and a chance to go out and enjoy Hawaii. A conference in this venue means that we certainly going to tempt people to skip a few sessions. This time we thought we give them an explicit opportunity without guilt to head off and enjoy.

… to enjoy authentic aloha spirit!

Indeed - There's lots of aloha spirit here. The convention center and the people are very Hawaiian. They've been very friendly and very accommodating. So far working with the convention center has been great.

There are two other interesting things this year that are novel. One is the EXPO and the poster session. There's been a few experiments with how those two should interact. This year, like in 2015 CVPR, we put the posters in the middle of the EXPO so the companies are in a ring surrounding the posters. The difference from this and 2015 is that we are much bigger now. In 2015, there were about 50 companies exhibiting. Now we have 125, and we have almost twice the number of posters so it's a huge arrangement. We hope it works. There is going to be a lot of flow. The expo is going to be open during the poster sessions.

"The industry spotlights"

Lastly we have what we call "the industry spotlights". On the EXPO floor, there is an area for this. It's just the size of 20x20 booth, but we have a full program throughout the entire conference during lunch and the poster sessions of companies giving 10 minute talks. The program is online, but every company who is in the EXPO is given an opportunity to participate in this. Most of them do so we have a running highlight for each company. You can see it on the program. I suggest you look at it and just drop by. It'll happen during lunch and also during the posters. It includes also a panel session bringing together people from a number of companies to talk about key issues in our field…

….which will be moderated by me…

Exactly!

What is your advice to a first-time attendee to CVPR? They will have thousands of presentations and lectures. What is your advice?

First, welcome to CVPR to all of the first-timers. We don't get the kind of growth we've seen without having a lot of people who have never been to CVPR each year. We have a conference survey that we do every year. We make this available to all attendees typically towards the end of the conference, and then it's open for weeks afterwards. I very strongly encourage everyone to fill out this survey because this is how we understand what people liked and what didn't go well so that we can plan the conference for next year. This is instrumental in us choosing a 4-day conference vs a 3-day conference for example, or choosing to go with invited speakers vs oral spotlights. All kinds of experiments that had been done over the past few years and new ideas, we use the survey to judge whether the people would like this to happen or continue.

Last year in the survey, about a quarter of the registered people responded. We would like to have that percentage to be higher, but according to that 45% of the people were new at CVPR. This might have been a bias in the sampling, but nevertheless, they're a huge percentage. We expect this year might be comparable or even higher at

least in terms of the response to the surveys. So, there's a lot of you out there. I think for all the new folks, it can be very overwhelming. I would advise you to treat it like a tasting menu.

"The poster sessions can be overwhelming"

Make sure that you get to a few of the oral sessions to see the high quality of our selective long oral presentations and some short orals. I would advise you to make sure that you drift around through the posters sessions and get a feel for that by highlighting the posters that you want to.

The poster sessions can be overwhelming. In each poster session, there will be about 112 posters. There's going to be 2 poster sessions each day, but posters will only be active on one morning or one afternoon session. I would advise you to figure out which posters and topic areas are of most interest to you and target those rather than getting lost in the crowd. There's all kinds of navigational help like an online floor plan, things like that so you can find the posters that you want to get to.

The EXPO should be quite a technology fest. Most of the major internet companies are here. Virtually every computer vision company is here. Many or most companies invested in machine learning are here. The EXPO grows in size and dynamism every year. It's a very exciting group of demos and presentations that companies are making. It's quite large so I would sample all of those.

"Don't forget the luau!"

Don't forget the luau! The luau is a Hawaiian party on Sunday night. This is the one major sponsored social event that comes your registration. It's down on the beach near the hotels, and it should be a great time. It will be a traditional Hawaiian luau with entertainment, native dancing, torch juggling, and things like that.

I would close by saying, please consider staying for the last workshop day. It's the day after the main conference. It'll be the 6th day, but we gave people a half day off on the fourth day so that they might be more likely in part to stay for that last day. Of course, Hawaii calls, and people want to go, but there's still a lot of great workshops and interesting talks happening all the way through the final closing of that workshop day. I would also encourage everyone to fill out that survey. It really is how we understand what the community wants, what they like, and what they didn't like. The new aspects of the conference will be represented on the survey, whether you like three parallel tracks or whether you want four or five parallel tracks.

Please fill that out. That's how we know whether you like the poster session, the EXPO layout, and things like that. That is how you make your opinion heard. The community has spoken very loudly on many of these issues with 95% voting in favor of the EXPO, invited speakers, and so on. You can always send an email, but the survey is data, and data speaks very loudly.

Modar Alaoui - Eyeris

CVPR DAILY
July 21-26 2017 Saturday

"We are hardware agnostic, camera sensor agnostic, gender and authenticity agnostic…"

Modar, please tell us about Eyeris.

We are a deep learning-based vision AI company. We do computer vision work for human understanding. Particularly, right now, for face analytics and emotion recognition. Because we believe that we can derive the richest source of information about human behavioural understanding through faces. And then some of our additional product roadmaps are to complement the face analytics and emotion recognition with body pose estimation, understanding, action recognition, activity recognition, etc.

Modar Alaoui, founder and CEO of Eyeris

Would you agree that this segment has already a significant number of actors working on this?

Yes, so there are a lot of companies that do a lot of different things. What makes us special, and I think that's kind of your question, is: number one, because we focus on face analytics and emotion recognition today, we provide the largest number of analytics or deliverables ever extracted from the face today. That includes seven emotions, age, gender, head pose, eye tracking, eye openness, gaze estimation, all of that good stuff, attention distraction, drowsiness, microsleep, etc., etc., etc. I mean, I may have just mentioned here like 15 or 20, you can imagine what the list is. So, not only that we do all of this in one single, all-in-one SDK or software, but we also have built our proprietary deep learning-based methodology that is primarily using convolutional neural networks for, obviously, image processing. But our methodology includes embeddability in mind. So when we built the algorithm we wanted to make sure that the algorithm does not only sit in a big desktop environment with a ton of GPUs. We want to literally put it onto a Raspberry Pi, and other low-process, low-memory, low-power devices. Embedded devices particularly.

And why are you specifically so proficient in that?

Because we started initially with hand-tuned features before even deep learning came along. And we were literally pushing ceiling as far as accuracy, as far as what we can do with all of the traditional machine learning techniques. When deep learning came along, we knew exactly what we were missing, and that the big leap is about the extra 5% of 3% or whatever, so you can have the extra accuracy under different lighting conditions, under different environments, or what we

call different uncontrolled environments in general. So the methodology that we created, which is proprietary, kind of bridges between the two, if you would. But that is dependent on where the algorithms needs to be deployed. If it needs to be deployed into some GPU with massive amounts of processing power and all of that, then we use deep learning specific algorithms. And if it needs to be deployed into a Raspberry Pi then we kind of leverage some sort of what I would call a hybrid between the two so that I can remain a little bit vague about our methodology.

Well, this requires a bit of a story now. How did the idea come to you in the first place, at such an early stage?

So, my background is in user behavioural measurements using technology. And early 2008, I launched this company that was doing gender and age group specific advertisement, or gender and age group targeted advertisement using computer vision. So using age and gender recognition from a private in-store TV network that was launched or deployed in airport shopping centres and things like that. So the idea behind this was basically putting out advertisements based on gender on airports when you are within 10 or 15 feet away from the display. So we needed more tools that are kind of cutting-edge for measuring the audience. And so we weren't satisfied with age and gender, and people counting, and dwell time, and all of that. We started looking into emotions, and we exhibited as the first emotion recognition company ever at CVPR in Portland in 2014, and we literally were among a dozen companies like Amazon and the-liked, and we were the only startup. So that's basically when the idea came along. We had a good prototype, we showed it, and then we wanted to pursue a software-only kind of play where we would provide everything that we learned plus everything that we started to learn about emotions and to combine everything into a single software. Not only we have everything now, but we also are hardware agnostic, camera sensor agnostic, gender and authenticity agnostic. We trained on over three million images and videos of data that we collected initially, because there was nothing publicly available. We spent the better part of two years collecting data regularly and we still do that every day at our lab in San Jose.

In one of the talks today I heard that most data that is used to train such systems is collected without the permission of the person, or even after receiving a denial for collecting that information. I am sure that you have an answer to that?

[We laugh] - of course! We hire between 120 and 150 people to our San Jose lab every single month. We have the largest dataset ever collected inside a vehicle with people's consent. We keep everyone's record. Not only record - I mean we have self-reported age, gender, ethnicity, background, height, weight, etc., from every person with their IDs attached to release forms, which is what we call them. And that gives us full authority to do really whatever we want to do with the data. And what we do with it, really, is not to post it out there on our website, but to train our algorithm on it and leverage it

as ground truth. And before training we of course have to label the data, and so on.

"The grey area is much cheaper, much easier... ... and scalable"

What about the segment in general, do such companies as I mentioned earlier exist?

They do exist, unfortunately. Unfortunately they do exist. There is a grey zone where people go out on the internet and collect one million images from Flickr of from people's profile pictures of Facebook, or from other social media. And those people, yes, they have not granted permission for this company to collect, but indirectly they somehow gave permission to the media platform to basically show their results or their images publicly on the internet to anybody that is interested in that profile or that image. And that is why I say that it is grey. So it's indirectly given to the platform owner, but it's not necessarily direct authorisation to analyse it or to even train on it. So my take on this is that it is great to the extent where the algorithms that are trained on this data remain for non-commercial purposes. And the second that some algorithm that was trained on this data becomes for commercial purposes, I believe that there is a little of a bridge there being crossed.

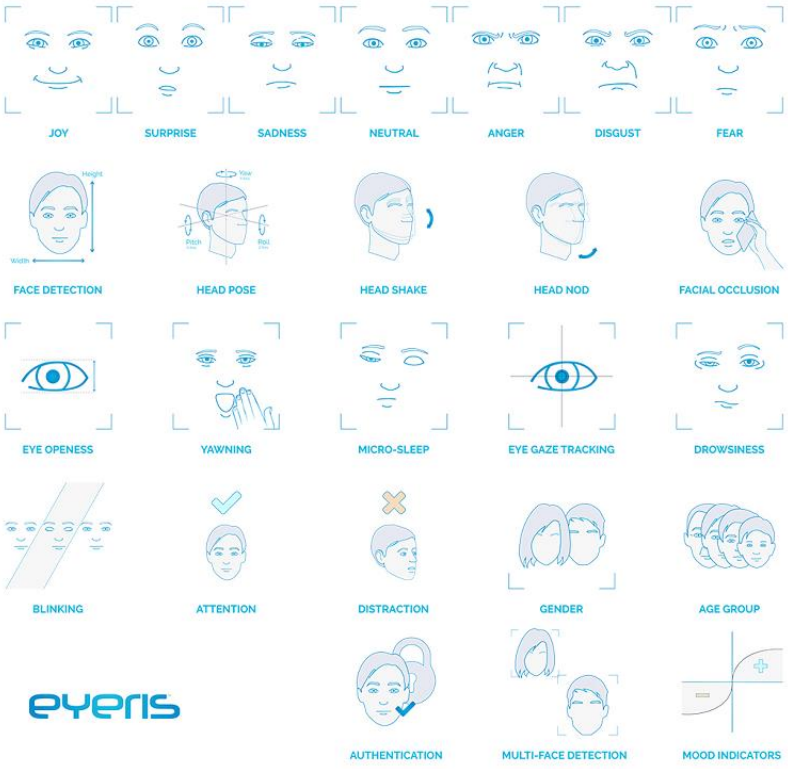
That means that this grey area is much cheaper and much easier...

... and scalable, yes.

That means that you forced on yourself to give up those three advantages in order to be sure that what you do is not in a grey area but in the white area.

That is right. So we did that for two reasons. The first is the same reason that you just mentioned, but also for the reason because the algorithm that we are producing is for a very hyper targeted environment and set-up. And so we go after two different types of verticals, one is autonomous vehicles and driver monitoring, or what we call occupants monitoring AI for highly automated vehicles. That includes the driver of course and all the other passengers. And then the second vertical is social robotics.

But I'm not going to talk too much about social robotics, because that's not really why we tried to collect all of that data. We collected the data inside the vehicle and so there is no publicly available data or no data out there on the internet today that you can leverage



A sample of Eyeris' all-in-one face reading AI deliverables for highly automated vehicles (HAVs)

to train an algorithm with different lighting conditions, non-uniform partial lighting, different head poses, person looking at different areas in the wind shield, different parts in the car, when the camera is placed in different positions, etc. We had to literally create all of that and also for the benefit of creating data that we can leverage to really do whatever we wanted to do with it and to train the algorithm for the purpose of selling it commercially.

"Scout around..."

What did you come to do at CVPR2017?

So, not only we are exhibiting here. I have a ton of meetings with a number of companies which I'm pleasantly surprised to see here, including startups or even larger companies. But so we are exhibiting and looking at some papers. Potentially we'll be voting for some of the papers that we are looking at, computer vision papers particularly, that we are looking out for behaviour understanding. And also my team is here to scout around, and meet and mingle with professionals in the field, potentially hire a couple of people that would be great for our mission and our company.

Does that mean that people can meet you and your team at the EXPO?

That's right - anytime during the next four days. Otherwise they can easily drop us an email and we can make something happen.

You can meet the wonderful team of Eyeris at booth 651.

Ilke Demir

CVPR DAILY
July 21-26 HONOLULU 2017 Saturday

Women in Computer Vision

Ilke Demir is a research scientist doing her Postdoc at Facebook.

Ilke, where did you study?

I did my PhD at Purdue University. I also did my Masters at Purdue in the Computer Science department, and I did my undergrad at the Middle East Technical University in Turkey.

Where was it more pleasant to study?

I don't know. My universities always had a special place for me.

Why did you choose this field?

Since my childhood, I always had that thing where I assembled and disassembled things. I was curious how things work. My father and me had tasks to solve things, build circuits, or see how things work… little experimental things.

Is he an engineer?

He is actually a helicopter pilot. Now he is in the private sector, but he used to be in the military.

So he's very much aware of technical things?

Yes - in order to have their license, they need to have a major from a university, so his background is in electrical engineering.

Since childhood, we used to do all those little activities. After that in middle school, I pretty much knew that

"Why should gender matter?"

I wanted to study computing tasks. I was really good at math. All my teachers used to say that I should have a better education and devote myself to math. I decided in high school to study computer science and electrical engineering. The Middle East Technical University is the best university in Turkey. There is a big entrance exam in Turkey. You enter all of the universities with this exam. Approximately 1.5 million students enter that every year. I had really good rank in the entering exams so I could choose any department I wanted in the university.

It seems like you were always encouraged to study this field.

That's right. In middle school and high school, we had math olympics and computer science olympics. In middle school, I won a bronze medal in math. After that I was super dedicated. In high school, my seniors, who are now doing awesome things in the Silicon Valley, taught me coding. In my country, we didn't have coding and computer classes in high school. We didn't have computer classes in my high school, at least in my time, so they taught me coding for those olympics. I felt like I could do anything with coding.

Why did you decide to come to America from Turkey?

My adviser in Purdue was the best in what I wanted to study, which was 3D reconstruction. At that time, I was more focused on robotic vision because of my internship in the Cowan Research Lab. I really wanted to be a robotics vision researcher. What my adviser in Purdue was doing at that time is reconstruction and appearance editing. That means having camera projectors to change a vase with flowers to a vase with checker board patterns and so on. It's visual, but it's real. I wanted to work with him on this project so I wrote to him to ask if I can work with him. I never got a reply, but that happens of course. Everyone applies everywhere. I said I am going to work with him anyway. The university already accept me so I knew I was going to Purdue. I did my whole PhD there.

How did you end up working with Facebook?

Together with my adviser at Purdue I was giving a course at SIGGRAPH on inverse procedural modeling. Ramesh Raskar is my supervisor on Facebook, he is new from the MIT Media Lab. He is the professor that everyone is running after. He came to our course at SIGGRAPH. He thought the idea was

applicable to the project he had and he liked my presentation - because it's a three hours and fifteen minutes long course, it's an afternoon course with two professors and myself.

It seems like everything has worked out well for you? Are you lucky to have everything work your way?

I don't think it's luck. I know that some things worked out because certain people happened to be in my life along the way. I worked hard on those things as well. As I told you, my advisor didn't accept me at first when I wanted to work with him. Then I took his class and showed him I was really interested in graphics. Then in my second semester, not even the first semester, he asked me to be his research assistant. You can overcome anything.

What is your current work?

We are trying to understand the world from satellite images. We started that work by extracting roads from satellite images. The big problem is 75% of the world is unmapped. By that I mean, if you want a parcel to be delivered to an address and the address that you put on Amazon for example is "Go to that market and behind that market, take the second street. Then you will see a temple there. It's the second house on the right." That is the address. I can't process what is behind the temple. You cannot have those addresses.

The main challenge that we are trying to solve is a automatic, generative approach to map all those places, to give everyone a unique street address that they can use for all of the location services, urban infrastructure, connectivity efforts, etc. Our work at CVPR was studying satellite images and processing them with a deep learning network to extract roads. Then we segment those roads to have an individual street. Then we name those streets automatically based on some distance fields, according to a four-field alphanumeric addressing scheme.

What is particularly challenging in doing that?

Every part had its own challenge. We experimented with different deep learning networks. We experimented with different region generation algorithms. Technically, it wasn't that challenging, but after that having someone use the system is the most challenging part. We can have the best address scheme, but if nobody is using it then why do you have it, right? We have some partnerships with some companies and government agencies in developing areas about using our addresses. Currently, we have really good feedback. For example, for some user studies, we have 25% reduction in arrival times. If you are using the traditional addresses, it takes you hours longer to find the place that you need.

[*laughs*] Even taxi drivers cannot find my address sometimes. Do you enjoy what you do?

To be a PhD student, you need to be dedicated. To dedicated, you need to love what you are doing. Even with this work that I'm doing with Facebook, it is kind of related to my PhD which is in inverse procedural modeling. I see the world as if I can proceduralize everything. This is a branch of it in which I proceduralize satellite images.

If you could proceduralize anything in the world, what would it be?

I've heard that question before! I would want to proceduralize evolution. That would include everything we have learned nowadays for humanity. It would include our learning process, it would include out growing process. It would include us, both semantically and geometrically. Just imagine that you are being proceduralized. You know how your finger grows from a cell. All of those are based on a grammar. So just imagine you that could express human evolution with one set of rules, with one procedure, one grammar. That would be so awesome.

Does it mean that everything is predetermined?

It can be a stochastic grammar, it doesn't need to be pre-determined. But it needs to be based on some rules.

So is there no power governing this world?

The power belongs to the person that is able to extract that grammar.

What do you think about us giving a voice to women in computer vision?

Ideally, the world should be equal. You shouldn't be focusing on women. Why not ask white males what they think, for example, right? But we live in a society where this is needed. I admire and appreciate what you are doing. I want to thank you for that. I know that when I speak, people hear me, but I

know that not it's not the same for all women in computer vision and computer science. I witnessed so many engineers and higher ups not having their voices heard, being overtalked, or not being looked in the eye because of their gender. That's why you need to continue this.

I heard two different approaches actually. I've heard that it might put women on a lower level. Others have said that you do need this positive discrimination otherwise you wouldn't be heard at all. What do you think?

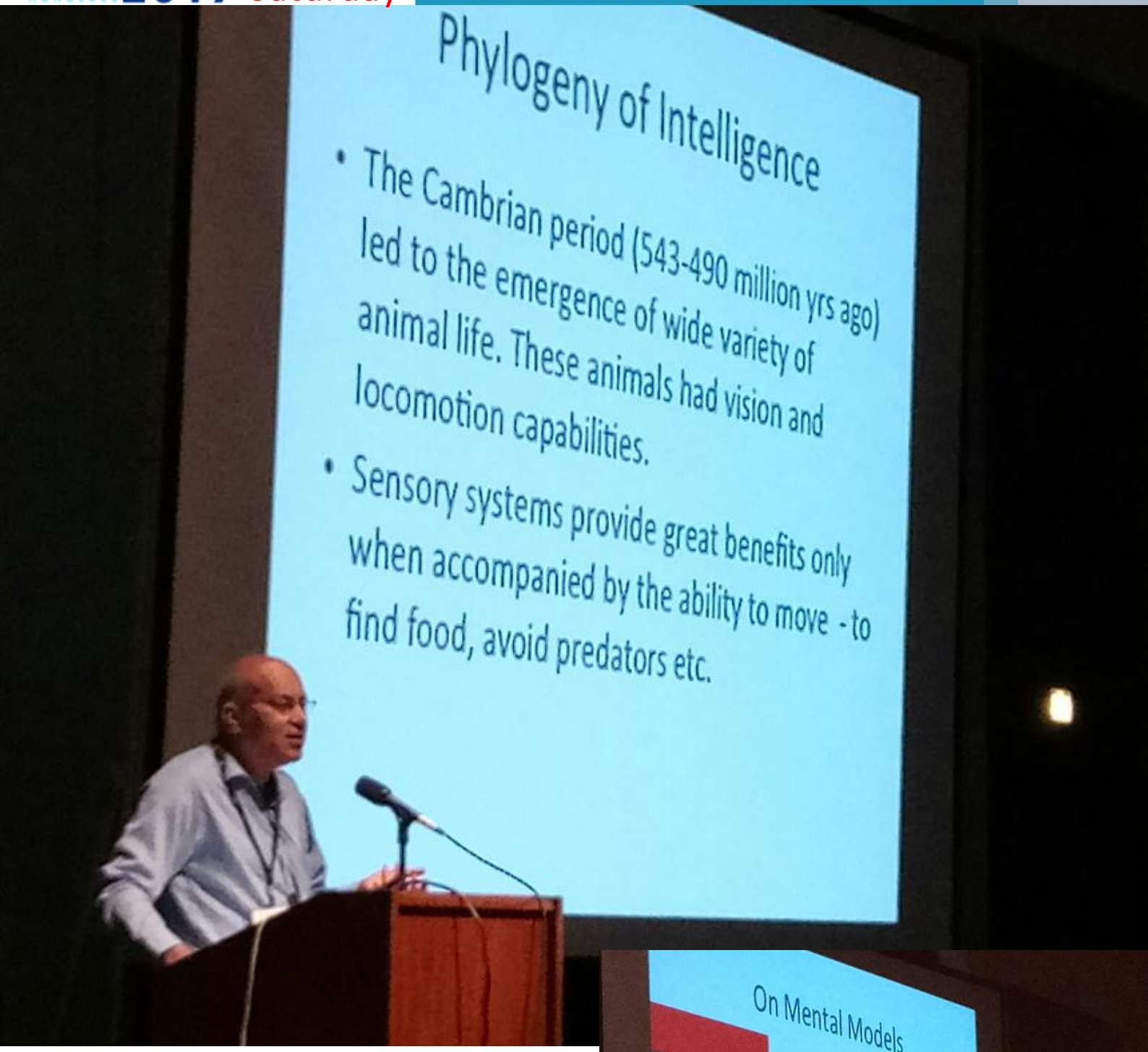
I think positive discrimination is still needed at least to encourage young girls. We keep seeing those numbers increasing, but we are still so far from the ideal equality. We need to do everything in our power. I'm not saying we should reduce the quality of the work because of gender. I'm not saying positive discrimination does not affects the overall quality of something. But if you can make those trade-offs, to empower a minority - which does not have to be women, it can be race, age, or any minority. But if you can empower a minority to have some part in the decision or in the process in the work environment then that is really something you should have.

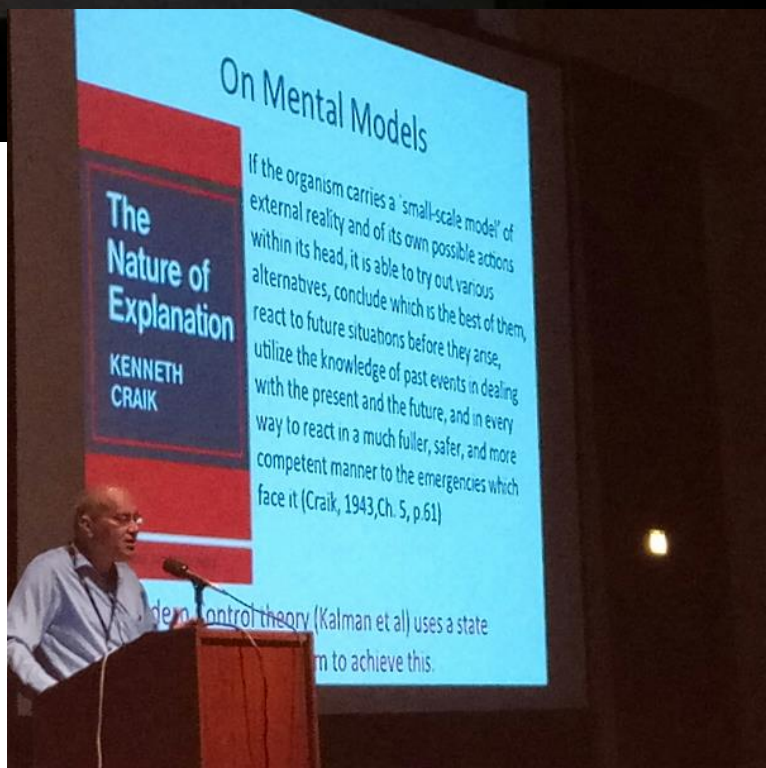
How can we improve the behaviors of people to help make this change?

I think everyone, in all interactions, should be seen as who they are. They should be abstracted from what they are, and everything before the conversation. It should be a human to human conversation. Your title, status, and achievements shouldn't matter when you have conversation in formal and informal settings. Those abstractions include everything; gender, class, race, etc.

Do you think men can do that when they are talking to a woman?

Why not? I don't see why not. Why should gender matter?





It is always passionating to follow a lecture given by Jitendra Malik. This year, at the Deep Learning for Robotic Vision workshop, he shared with us great quotes, good humor and new insights from his recent work with Google.

Silvia Vinyes

CVPR DAILY July 21-26 2017 Honolulu Saturday

Deep Learning for Domain-Specific Action Recognition in Tennis

Silvia Vinyes is starting the fourth year of her PhD at the **Imperial College London**. Yesterday she presented her poster at the **Computer Vision in Sports** workshop.

"In the dataset there were professional and amateur players, and the model performed different for each group"

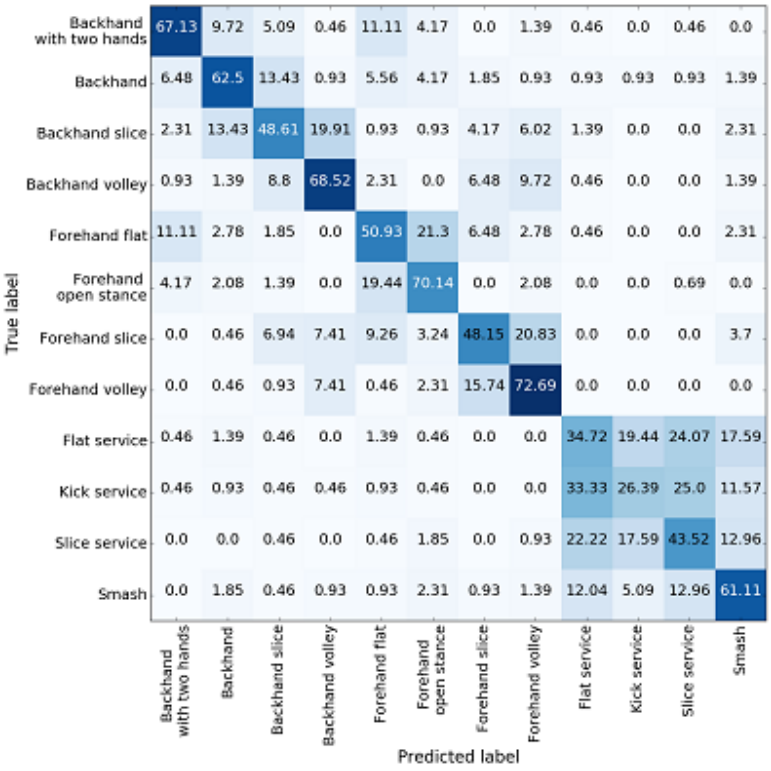
Silvia's work is about a method that she has developed for action recognition in tennis. What she and her co-authors are trying to do is recognise the fine-grained action of a tennis player. This includes for example specific types of serves - flat serve, kick serve - or different types of backhand or forehand. The idea is that "*in the future, this information can be used to build models of tennis*", Silvia explains.

In current models of tennis, you usually have only the position of the player and the position of the ball, and then you build a temporal model from that. Silvia thinks that if you know the specific action that a player is performing you can build an even more accurate model. Additionally, when you see a match then maybe you want to know which actions a player is weaker in or stronger, or which kind of strategies he uses. In that case, you would want to make a difference between the first and second serve, and see which kind of action he is performing. By automatically collecting the type of action, instead of doing everything manually, even more data for these kind of tasks can be collected.

There are also some challenges that Silvia told us about. One difficulty is that there are many things that change: the place from where you see the player can change, the type of surface where they are playing on can change, or the illumination in the video can change. Even how a player performs an action changes because the movement they do is not exact. Using this information, you could even think of another application, Silvia says: if you know the way of serving of a specific player and he changes the way of serving maybe you can see that something is going on, like an injury.

The way Silvia and her co-authors approached the problem is to use a deep neural network that was trained on an independent dataset to extract some features. "So we let the network decide which features are important in the image", she tells us, and then they used these features to tell which action is being played, using another neural network.

The first network they used is the inception network from Google. On top of this, they built a three-layer LSTM. Silvia tells us about an interesting thing they found after training the model: in the dataset they used there were professional and amateur players, and the model performed different for each group. Also, when the model was trained on one group, it performed good when tested on this group, but worse on the other. She says it will be interesting to see why it's performing better in one or the other.



The next steps Silvia sees for their work is to test the approach on other, more challenging datasets. In the dataset they used the background and illumination are changing but it's not actually in a real world setting since the players were not actually playing a match. At the same time the viewpoint in these videos was not changing that much so it would be interesting to see from different viewpoints, if they can achieve the same results. "*In an ideal world what I would like to have is more data, but for this type of problem the data is restricted*", Silvia says, and adds that "*this is why it's interesting to work in this type of setting*". The data is restricted because it is a very specific problem, she tells us. If you're trying for example to detect which sport is being played, there are datasets with a lot of data which are available because you can just go to YouTube and then collect videos of different ports. But settings of different fine-grained tennis actions the data is much more restricted. One of the things Silvia is particularly happy about is that they used a dataset that is publicly available, since in the area of sports action recognition most of the time people use their own data and its hard to compare the results. Silvia hopes this encourages people to use the same dataset and compare to their approach.

Roey Mechrez

CVPR DAILY
July 21-26 HONOLULU 2017 Saturday

Template Matching With Deformable Diversity Similarity

Roey Mechrez is a PhD candidate at the Technion of Haifa, in Israel

"A trophy held by the One Direction band is correctly detected in the target image, regardless of the occlusion: as I said, we only need one direction..."

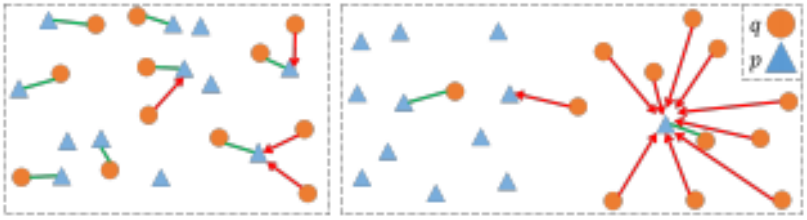
Roey Mechrez and Itamar Talmi will be presenting today their work "**Template Matching With Deformable Diversity Similarity**": a novel measure for image similarity named Deformable Diversity Similarity (DDIS) – based on the diversity of feature matches between images.

Methods for measuring the similarity between images are a key component in many computer vision applications such as template matching, symmetry detection, tracking, counting objects and image retrieval.

The key contribution of this work is a novel similarity measure that is robust to complex deformations, significant background clutter, and occlusions. The authors demonstrated the use of DDIS for template matching and showed that it outperforms the previous state-of-the-art in its detection accuracy while improving computational complexity.

The main challenge of this development, according to Roey, was to give robustness to the model in front of all kinds of different template matching problems: *"It is hard to model background clutter, deformations and occlusions, and yet, we want to have a similarity measure which is able to detect the best match. That is why our method relies on both local appearance and geometric information that jointly lead to a powerful approach for similarity. It is*

*based on two properties of the Nearest Neighbor (NN) field of matches between points in two compared images. The first is that the **diversity** of Nearest Neighbor matches forms a strong cue for similarity. The second key idea behind DDIS is to explicitly consider the **deformation** implied by the Nearest Neighbor field. The idea of allowing deformations while accounting for them in the matching measure is highly advantageous for similarity.*"



(a) similar distributions (b) different distributions

The team has continued working on the problem of image similarity, pursuing new research directions for DDIS. Roey concludes with a story: "*Comparing to the previous state-of-the-art, called Best-Buddies Similarity, we claim that we need only one direction of the nearest neighbor field and there is no need for both directions. To make that point, one of the figures in our paper shows a trophy held by the One Direction band which is correctly detected in the target image, regardless of the occlusion: as I said, we only need one direction...*"

Do you want to learn more? Roey and Itamar will present their work today (Saturday) at 13:30 - Kalākaua Ballroom C

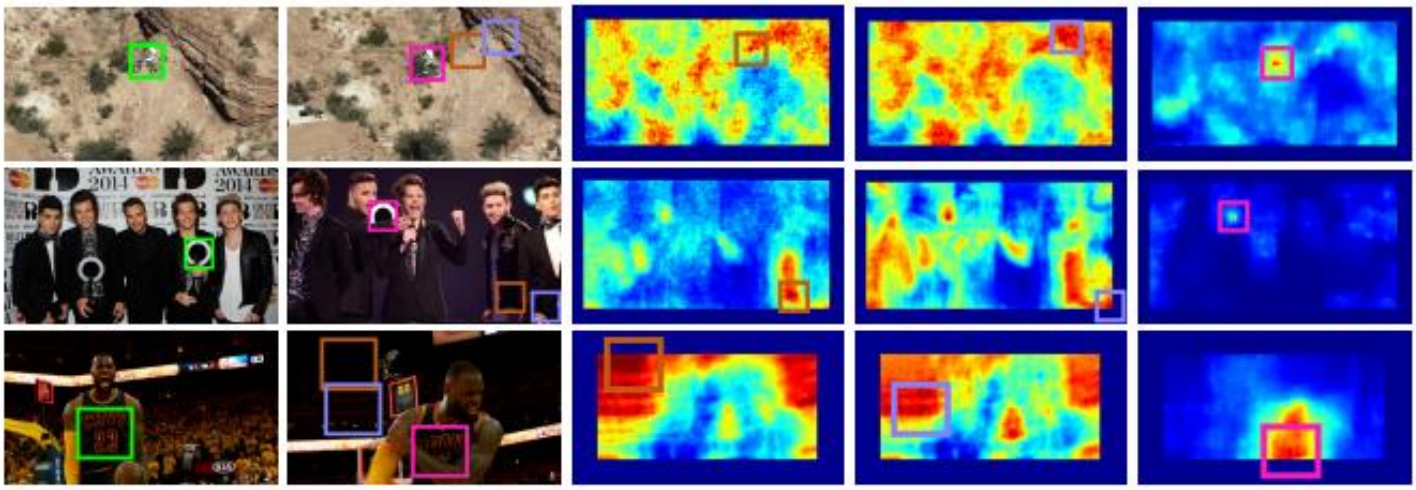
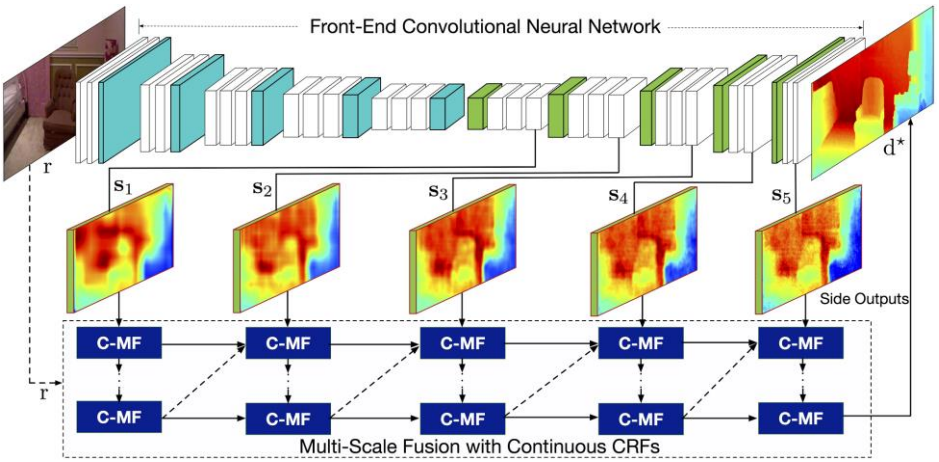


Figure 8: **Qualitative assessment**: The template marked in green (a), is detected in the target image (b) using three Template Matching methods: BBS, DIS and DDIS (all using the RGB features). (c-e) The corresponding detection likelihood maps show that DDIS yields more peaked maps that more robustly identify the template. Going over the rows from top to bottom: (1) BBS prefers a target location where the background matches the template over the location where the motorcycle is at. This happens because the motorcycle deforms and hence there are few bi-directional correspondences between its template appearance and target appearance. DIS and DDIS use more information – they consider all the one-directional correspondences. Therefore, they locate the motorcycle correctly. (2) The trophy won by the band *One Direction* is fully seen in the template, but occluded in the target. Nonetheless, DDIS finds it (as Section 4.1 said, we only need one direction...). (3) Complex deformations together with occlusion confuse both DIS and BBS, but not DDIS.

CVPR DAILY July 21-26 2017 Saturday

Multi-Scale Continuous CRFs as Sequential Deep Networks for Monocular Depth Estimation

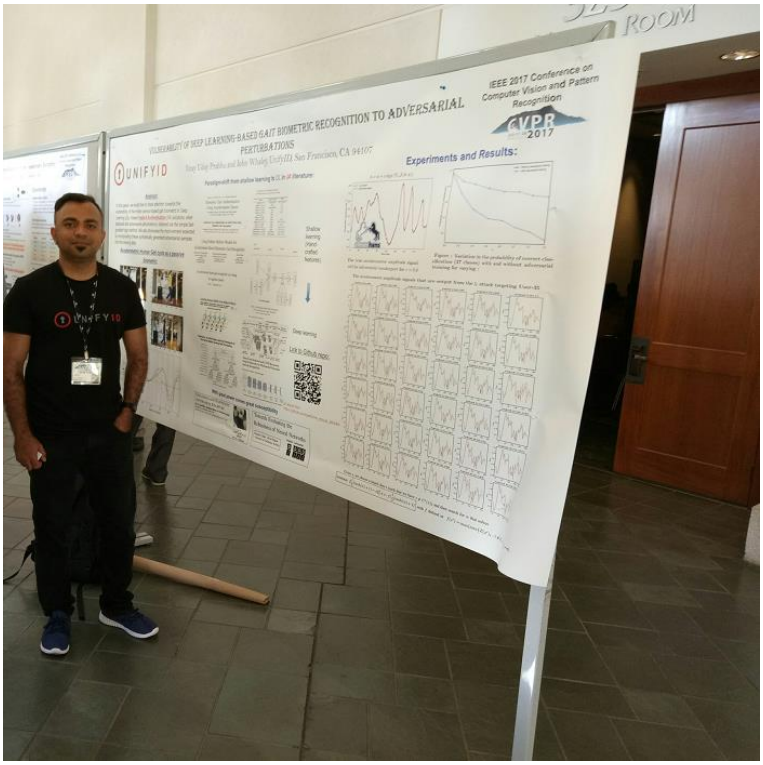
"It's also useful for other computer vision problems, involving continuous variables"



Dan Xu is a PhD candidate at the **University of Trento, Italy,** and he is also holding a research assistant position at the **Chinese University of Hong Kong**. He is presenting today his work about **Multi-Scale Continuous CRFs as Sequential Deep Networks for Monocular Depth Estimation**. The co-authors of this paper are **Elisa Ricci** who is working at **FBK**, a research institute in Italy; **Wanli Ouyang**, a research assistant professor at the **Chinese University of Hong Kong**; and **Xiaogang Wang**, who is associate professor at **Chinese University of Hong Kong**. Last author is **Nicu Sebe**, who is a full professor at the **University of Trento**.

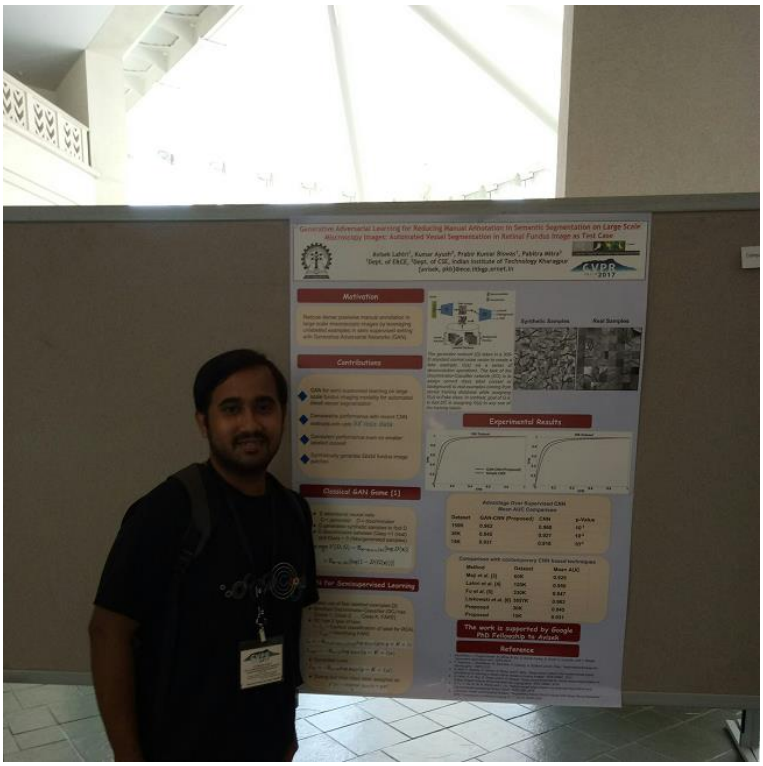
Their paper is about developing an end-to-end deep learning system to recover 3D depth information from a single RGB image. There are two novel aspects of this work, Dan tells us. First, previous work on this topic used graphical models, namely **CRFs**, but they only consider a single scale. "*But the multi-scale information has been proven very effective for a lot of computer vision problems*", Dan explains, "*so we thought in our task, the multi-scale information would also be useful*". So they considered to use the multi-scale information derived from the deeper networks, and to combine this information they propose a multi-scale (instead of just a single-scale) CRF. The second novelty is that they demonstrated the effectiveness of their framework when plugging the proposed multi-scale CRFs into different front-end convolutional neural networks. They show that for other tasks, if they involve continuous variables, the model and implementation of Dan Xu's method can be used as well. "*So this is not only useful for the monocular depth estimation, it's also useful for other computer vision problems, involving continuous variables*", Dan says, for example in crowd counting system, where the number of people in a crowd has to be counted.

If you want to learn more, you can attend Dan's spotlight presentation today (Saturday) at 9:20.



At the workshop The Bright and Dark Sides of Computer Vision: Challenges and Opportunities for Privacy and Security, Vinay Uday Prabhu presented: Vulnerability of Deep Learning Based Gait Biometric Recognition to Adversarial Perturbations. These are Adversarial Perturbations in a non-computer vision setting, accelerometric data and sensor data.

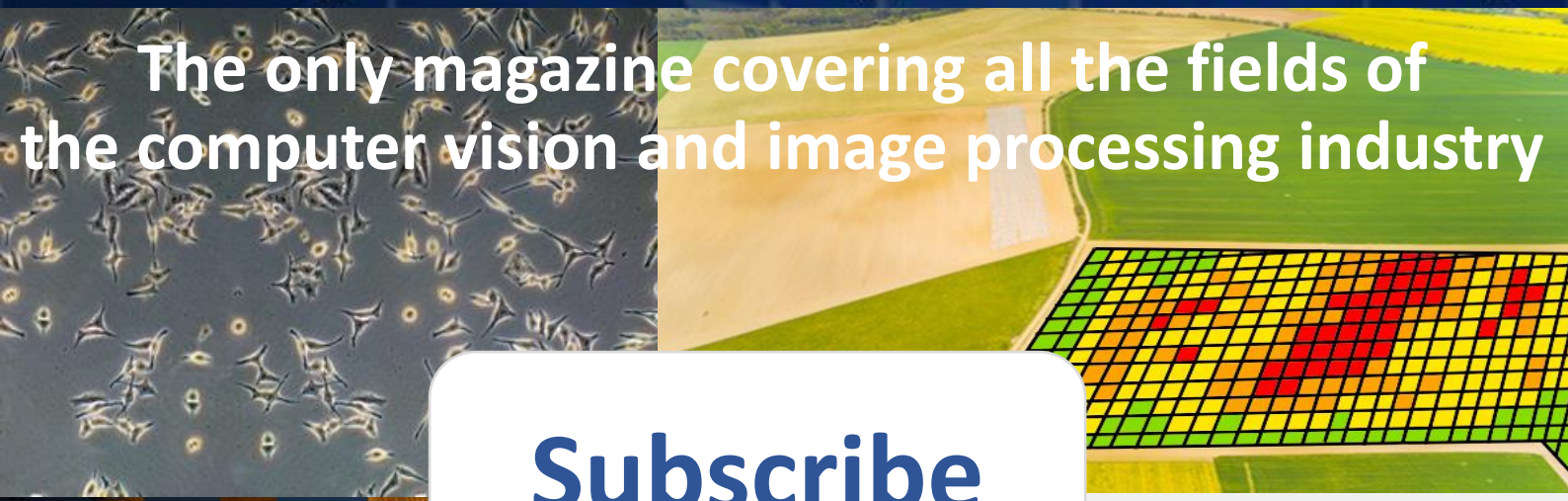
At the workshop Computer Vision for Microscopy Image Analysis, Avisek Lahiri presented: Generative Adversarial Learning for Reducing Manual Annotation in Semantic Segmentation on Large Scale Microscopy Images. That is, how to reduce labor in computer vision annotations using adversarial GAN learning



Improve your vision with

Computer Vision News

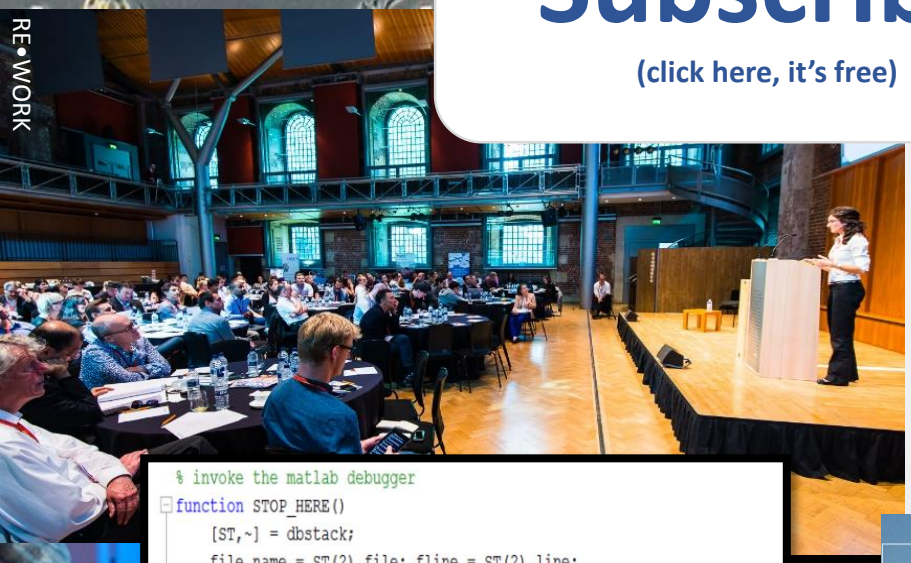
The Magazine Of The Algorithm Community

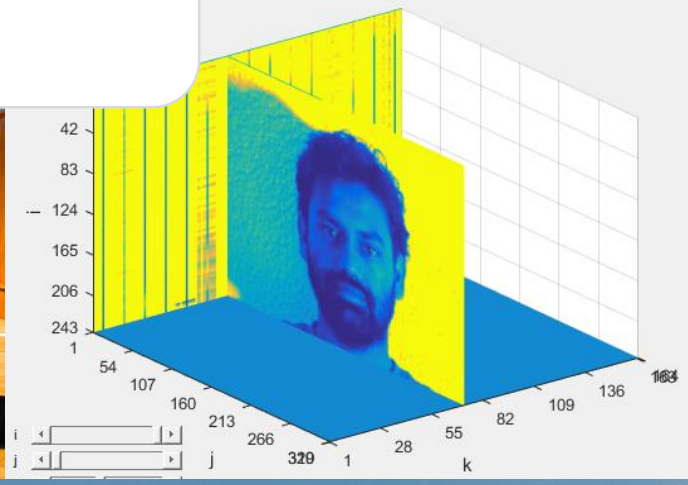


The only magazine covering all the fields of the computer vision and image processing industry

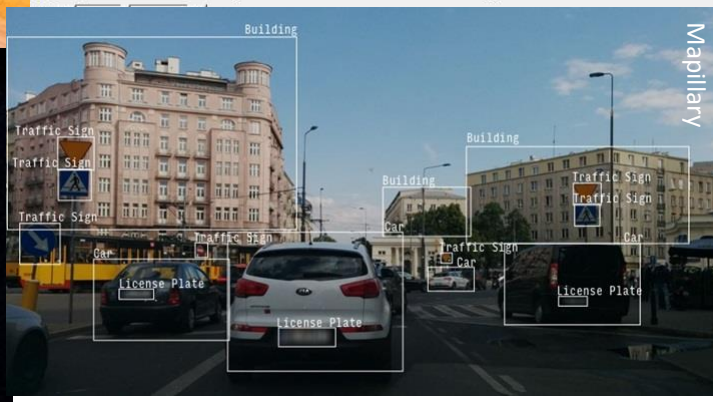
Subscribe

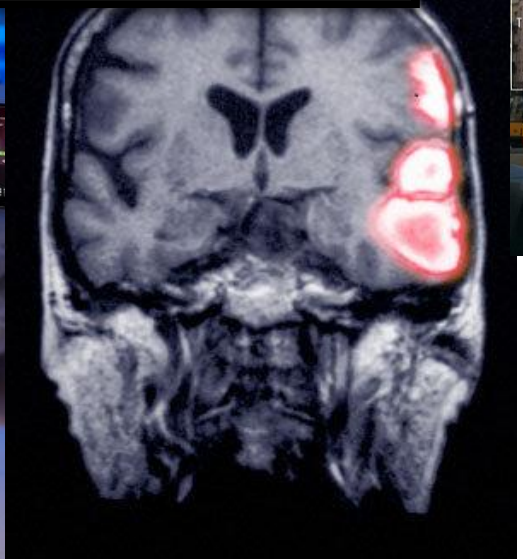
(click here, it's free)






```
% invoke the matlab debugger
function STOP_HERE()
    [ST,~] = dbstack;
    file_name = ST(2).file; fline = ST(2).line;
    stop_str = ['dbstop in ' file_name ' at ' num2str(fline+1)];
    eval(stop_str)
```





A publication by

RSIP VISION