FACULTY OF ELECTRICAL ENGINEERING
Doc. Mgr. ONDŘEJ CHUM, Ph.D.
ASSOCIATE PROFESSOR

October 16, 2020, Prague

Report on the PhD Thesis of Jitka Kostková "Advanced Moment Invariants for Pattern Matching"

The thesis addresses the problem of image and vector-field description that is invariant to various image transformations. Two types of transformations are studied in the thesis. First, geometric transformations of a vector field, namely rotation and an affine transformation are considered. Second part of the thesis addresses invariants to blur by Gaussian kernel applied to image (color channel) intensities or other functions, such as color histograms. Transformation-invariant descriptors are essential building blocks of many applications in computer vision, ranging from recognition, image retrieval to scene matching and 3D reconstruction. The thesis theoretically studies moment-based invariants. Novel constructions of invariants are proposed, these invariants cover image transformations that were not covered previously. The methods are rigorously described and analyzed. The results presented in the thesis clearly broaden our knowledge and understanding of the moment-based invariants. On the downside, the thesis does not provide convincing and realistic experimental evaluation. In the motivation section, page 4 Figure 1, the enormous number of images taken by mobile devices every year is shown. Yet, virtually all experiments in the thesis are synthetic and performed on a single or a few image instances.

The thesis is submitted as a collection of four journal articles. These articles were published in established journals, namely IEEE Transactions of Pattern Analysis and Machine Intelligence (IF 17.861), two articles in Pattern Recognition (IF 7.196), and Multidimensional Systems and Signal Processing (IF 2.38). Additional publications co-authored by Jitka Kostková include a number of international conferences: 2x CAIP 2019 (CORE B), SCIA 2019, ICIP 2018 (CORE B), and one journal publication in Signal Processing 2017 (IF 4.384). The publication tack of Jitka Kostková satisfies the expectations on publications behind a solid Ph.D. thesis.

The thesis is divided into two parts. The first introductory part provides brief motivation, review of prior work and summarizes the contributions of the thesis. This part of the thesis is written in good English. The second part is a collection of four journal publications. Each of the four publications and its contributions are addressed individually in the following paragraphs.

[A] Yang, Kostková, Flusser, Suk, Bujack: **"Rotation invariants of vector fields from orthogonal moments"** in Pattern Recognition 2018. According to Google Scholar, this paper has been cited 18 times. Rotation invariants of vector fields (specifically invariants to total rotation of the vector field, where the vectors rotate jointly with the support) were introduced. The invariants are based on orthogonal moments.

The experiment on the hair image uses image gradients as the vector field. It would be interesting to compare the proposed matching method with some standard image matching techniques – some of these even use image gradients to describe the patch – eg. SIFT. How exactly were the synthetic circular templates extracted? Was the rotated patch re-sampled to have axis-aligned grid? Did the central point / pixel remain exactly central for the patch?

[B] Kostková, Suk, and Flusser: **"Affine invariants of vector fields"** in IEEE TPAMI 2019. Despite this publication being very recent, it has already been cited (according to Google Scholar). This work addresses invariants to more general geometric transformation – the affine transformation. Novel moment-based invariants were derived. An algorithm to remove redundant (even polynomially dependent) invariants was proposed and executed to deliver irreducible set of invariants up to certain weight.

There are alternative approaches to direct computation of invariants from the (transformed) function. In particular, transformation equivariant (often called co-variant) coordinate system is constructed, so that any measurement in this coordinate system is an invariant. This approach dominates most of the computer vison problems where local features are used, such as structure from motion, as it was shown to outperform moment-based invariants. Would it be possible to apply an approach based on co-variant coordinate systems to vector fields?

[C] Kostková, Flusser, Lébl, and Pedone: **"Handling Gaussian blur without deconvolution"** in Pattern Recognition 2020. Novel invariants to Gaussian blur, with arbitrary shape of the Gaussian kernel, are derived in Fourier domain by a non-linear projection operator. Direct computation of the invariants in the image domain was shown to be equivalent, numerically stable, and does not require explicit construction of the projection. Further, it was shown that the proposed invariants to Gaussian blur can, due to its shape generality, be combined with up to affine geometric moment invariants.

One of the experiments is performed over rectified face patches. I presume that the method could be used to describe patches extracted from local image features as well. Such features are commonly blobs, corners, or saddle points. Would there be any difference between these types of features?

It would be interesting to see how well the proposed invariants perform if no blur is present – compared to state of the art approaches. Is the invariance to blur an additional desired property, or is it a performance trade off? This analysis probably cannot be shown on a subset of 38 face patches.

In the conclusions, there is a paragraph on possible future directions: "In a future work, it would be interesting to couple the proposed blur-invariant representation with the CNNs in order to make the CNNs blur-invariant without any data augmentation. …". This, indeed, would be an interesting direction. Is there any progress in such an application of the work?

[D] Kostková and Flusser: **"Robust multivariate density estimation under Gaussian noise"** in Multidimensional Systems and Signal Processing 2020. The key idea of this part is that additive normally distributed noise in RGB channel intensities translates to Gaussian blur in the color histogram. The technical contribution seems to be identical to [C] (equation (16) in [C], (23) in [D]). The target application is image retrieval based on color histogram. This application is supported by three references form 1990's.

In the experiments, for an image, 100 independent realizations of additive Gaussian noise are used to generate a set of 100 invariant descriptions. Then, an average of the invariants is compared with the invariants computed over the original image. It is concluded that "We can see that almost all errors are reasonably low." Is this enough to show that the invariants are good?

I have two further questions related to the whole thesis. 1) Different number of moments was used for each experiment. How was the number of moments (moments up to which degree to use) selected? Were any measures taken to avoid overfitting to data of a particular experiment? 2) What kind of similarity measure was used in the experiments? Do all invariants provide the same range of values or is some form of normalization needed?

Overall, the thesis introduces good and rigorous theoretical contributions. The results were previously published in prestigious journals; the work is being cited by other researchers. The author of the thesis proved to have an ability to perform research and to achieve scientific results. I do recommend the thesis for presentation with the aim of receiving a Ph.D. degree.

Doc. Mgr. Ondřej Chum, Ph.D.
Associate professor