

Óbuda University

PhD Thesis
Summary



Multi-Directional Image Projections with Fixed Resolution for Object Recognition and Matching

Gábor Kertész

Supervisors:

Dr. Zoltán Vámosy

Dr. habil. Sándor Szénási

**Doctoral School of Applied Informatics and
Applied Mathematics**

Budapest, 2019.

1 Background of the Research

The use of computer vision and image processing in traffic analysis and control has shown significant growth in recent years. In addition to traditional solutions (such as traffic counting, accident detection), multi-camera applications based on vehicle identification and tracking have also emerged [Sanchez et al., 2011].

It is not always possible to identify a vehicle by its license plate number. In addition to weather conditions, the lack of natural light can cause poor visibility. For example, reading a remote identifier on low-quality images of cameras in artificially lit tunnels can be challenging, as a result of noise, lack of color information and low resolution.

The task falls within the scope of object recognition within computer vision. The task is object detection when visual identification of an object of known appearance is required. Object classification is specified as defining the type of an unknown, segmented object.

In case the task is re-recognizing or re-identifying a previously seen object, the task is referred to as instance recognition or object matching [Szeliski, 2010].

Over past decades, the problem of pairing objects has been approached in many ways. There are many examples in the literature of solutions using keypoint based descriptors [Shi and Tomasi, 1993]; [Lindeberg, 1998]; [Schmid, Mohr, and Bauckhage, 2000]; [Lowe et al., 1999]; [Bay et al., 2008], specifically for vehicle recognition [Guo, Hsu, et al., 2007]; [Guo, Rao, et al., 2008]; [Hou, Wang, and Qin, 2009]; [Choe, M. W. Lee, and Haering, 2010] as well.

If the image is of poor quality and low resolution, it is not advisable to experiment with high-level keypoint-based solutions. In this case, pixel-intensity-based solutions may be used, such as template matching [Brunelli, 2009], Haar-like features [Oren et al., 1997]; [Papageorgiou, Oren, and Poggio, 1998]; [Viola and Jones, 2001a]; [Viola and Jones, 2001b]; [Yuan and Sclaroff, 2009]; [Rios-Cabrera, Tuytelaars, and Van Gool, 2012], or using projection features [Betke, Haritaoglu, and Davis, 2000]; [Liu, Collins, and Tsin, 2002]; [S. Lee, Liu, and Collins, 2007]; [Shan, Sawhney, and Kumar, 2005].

In the publication by [Jelača et al., 2013], a complex model was presented that is capable to identify vehicles with low error rate on distant, non-overlapping camera images. The method is based on a four-dimensional projection signature consisting of the horizontal, vertical, diagonal and antidiagonal projections. The projection signatures are comparable and their values can be used to calculate the degree of similarity between objects.

The dimension number of the projection signature can be increased by introducing multiple projection directions. The Radon Transformation [Radon, 1917], which became known primarily for its inverse formula [Deans, 1983], can be used to map to multi-directional image projections. A related mapping is the Trace transformation [Kadyrov and Petrou, 2001], which gives a generalized form of the Radon. The Hough transformation [Hough, 1962] is a method for detecting lines and shapes; it uses rotation-based mapping similar to the Radon [Deans, 1983]; [Ginkel, Hendriks, and Vliet, 2004].

The explosive growth of image classification efficiency in recent years [Krizhevsky, Sutskever, and G. E. Hinton, 2012] which is the result of machine learning; in particular the rise of deep learning. [Bengio, 2012a]; [Y. LeCun, Bengio, and G. Hinton, 2015]; [Schmidhuber, 2015]; [Goodfellow et al., 2016].

Modern image object recognition is based on convolutional neural networks¹, [Y. LeCun, Boser, et al., 1989]; [Y. LeCun, Bengio, et al., 1995], which is highly effective at large neuron and layer numbers [Y. A. LeCun et al., 2012]; [Bengio, 2012b]; [Szegedy et al., 2015]; [Karpthy, Li, and Johnson, 2017]. The large number of elements and the associated large number of training samples require considerable computational capacity. *Deep learning* methods have become widespread because of their parallelization, graphical processors can reduce the necessary training time to a fraction [Raina, Madhavan, and Ng, 2009].

[Bromley et al., 1994] used an architecture based on convolutional neural networks to pair image objects called the Siamese neural network². The method has been successfully used for validation of handwritten signatures.

Siamese architectures are used effectively for face-identification [Chopra, Hadsell, and Y. LeCun, 2005]; [Taigman, Yaniv and Yang, Ming and Ranzato, Marc'Aurelio and Wolf, Lior, 2014]; [Schroff, Kalenichenko, and Philbin, 2015]; they can also be used for *one-shot learning* problems, where few training samples are available [Oriol Vinyals and Charles Blundell and Timothy P. Lillicrap and Koray Kavukcuoglu and Daan Wierstra, 2016]; [Koch, Zemel, and Salakhutdinov, 2015].

2 Research Goals

The primary goal of my research is to analyze the applicability of multi-directional projection descriptors for object matching. Four-dimensional projection signatures are well suited for object matching. It can be assumed that the accuracy of the method can be improved by introducing multiple projection directions.

In the case of Radon transformation, the length of each projection depends on the size of the input image and the direction of projecting. If the similarity score is determined from the best fit of the projections of different lengths, then it may falsely indicate high similarity between different observations.

My goal is to develop a multi-directional image projection method that has a fixed, predetermined length. This would result in a mapping independent of the angle of rotation and the size of the image.

Investigating the parallelization capabilities of this method is also an important task, as modern computer architectures provide a high degree of support for data parallel execution, even for real-time processing, depending on the task. Typically, in the case of multidimensional arrays and matrices, the multiprocessors on the graphics accelerator can be used for parallel execution.

My further goal is to use the developed projection method for image matching, comparing it with other projection methods in terms of performance, computational complexity and memory cost.

In the next phase of my research, I examine machine learning-based methods for object matching and analyze the applicability of Siamese models based on multi-di-

¹Convolutional Neural Network, CNN

²Siamese Neural Network, SNN

rectional projection descriptors. The significance of different projection directions may vary depending on the type of problem and camera settings. It can also be assumed that similarity measurement can be performed based on different features of the parameter space.

A comprehensive experiment is necessary to determine the efficiency of neural networks based on projection features. To achieve this, it is necessary to develop a method for generating neural architectures for inputs with different types and sizes.

Thereafter, training and evaluating the large number of neural models is a computationally demanding problem that may be appropriately solved in a distributed environment. This problem is to design the training of multiple independent models where parallel efficiency is maximized by structure and scheduling.

Therefore, my goal is to design and implement a complex experiment on multi-directional image-projection-based object matching using neural networks. As part of this, I have to solve the problem of generating neural network architectures, distributed training and evaluation of the models.

3 Methods of Investigation

3.1 Multi-directional Image Projections

The length of the projection section is determined by the size of the input image and the projection angle. The projection length given by trigonometric functions using angle α is $\cos \alpha N + \sin \alpha N$, where N denotes the width of the squared image. This section can be divided into S bins with equal length and can be resolved independently of α angle and N length.

After defining the bins, the value of each examined pixel is added to the value of the covered bins in proportion to the overlap. The different number of affected bins should be handled separately. With exactly one affected bin, the intensity value is added to the bin entirely. With exactly two involved bins, the values are increased in proportion to the overlap. In the case of multiple affected bins, the fully covered ones are increased entirely and the partially covered ones are increased in proportion to the projection. After normalization with the affected element numbers, the scale becomes uniform.

In a data parallel implementation, it is necessary to consider the architectural features of the multiprocessor environment. In GPGPU³ programming, it is important to minimize the transfer between the host and the memory of the graphical accelerator to reduce the *overhead* of transfer time.

The extensive utilization of the GPU⁴ memory architecture is similarly important, as processors access local *on-chip* memory more efficiently than higher-capacity global memory.

Eliminating common variables between different threads run by multiprocessors is an important design step: using common variables can cause computational errors as a result of the race condition. Classic synchronization-based handling of the race condition leads to a decrease in efficiency; therefore, the restructuring of the

³General-Purpose Computing on Graphics Processing Units

⁴Graphics processing units

algorithm is a better design choice. At the accumulation of the results, reduction or atomic operations can be applied.

To determine the similarity factor of projection functions, the Pearson correlation coefficient is calculated. Based on the obtained correlation values, the Euclidean norm is normalized with the dimension number of the projection signature, which gives the similarity score.

To determine the effectiveness of the classification, a threshold technique can be used, where the similarity values calculated for the real pairs in the dataset exceed exactly 50 or 80 percent. After calculating such thresholds, it is necessary to examine the proportion of false pairs above the threshold, which are, false positives.

3.2 Siamese Convolutional Neural Networks

Siamese architected convolutional neural networks consist of two identical convolutional heads in which, in addition to the architecture, the weight and bias values of the parameters are equal (Figure 3.1).

Convolutional neural networks have a well-known layered architecture. Feature extraction is based on convolutional layers, and then occasionally pooling layers are used to reduce the representation size. Prior to the output layer, conventional fully connected layers are often used. In these network architectures, a variety of design patterns can be observed, from kernel or window sizes, through the number of filters to various hyperparameters.

The method developed to generate convolutional network architectures is based on three steps:

1. defining the maximum kernel and window sizes based on the size of the input matrix and the number of convolutional layers;
2. generation of convolutional and pooling layer pairs, having a valid representation while keeping the maximum memory cost;
3. collecting a certain number of architectures and optimization of the training batch size for maximum memory utilization.

The *Master/Worker* design pattern is a convenient structure for training the generated models in a distributed environment, as the processes are independent of each other, only the input training data are common.

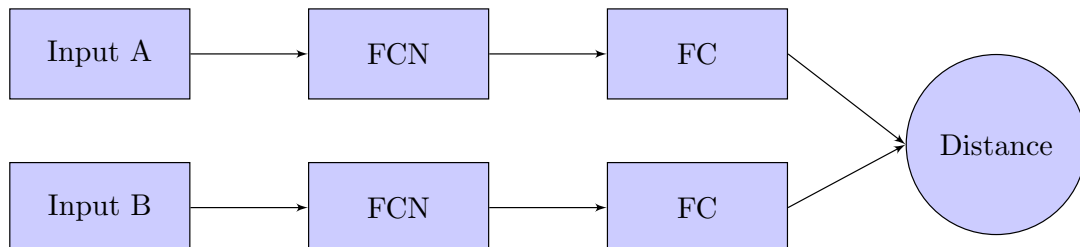


Figure 3.1: The structure of the two-headed Siamese Neural Network. The fully-convolutional (FCN) layers are followed by fully-connected (FC) layers. The parameters of the heads are identical, the outputs are multi-dimensional vectors. The measured distance of the output vectors give the similarity of the inputs.

A characteristic property of the pattern is that load balancing is automatic, and *Worker* units receive tasks from the work queue on request. Parallel efficiency can be increased by scheduling, for example, heuristics based on handing out tasks in descending order of the execution times. While execution times are not known in advance, they are estimated from the parameters of the architectures.

The matching performance of models can be measured using *one-shot* classification, which takes pairing accuracy into consideration instead of examining measured semantic distance values. In the latter case, similarity prediction is done by thresholding the distance between the feature vectors, which appears as the output of the Siamese network. In contrast, in the case of *one-shot* classification, the reference object is compared to several objects belonging to different instances, similarly as in real-life applications. Classification is correct if the distance between the reference and its real counterpart is minimal compared to other objects.

The models are also comparable in terms of memory cost, which is well approximated by the number of weight and bias parameters in the model. Multi-objective optimization can provide optimal models based on classification accuracy and model memory cost.

If a model m_1 is better than model m_2 in terms of predicted performance and in estimated memory usage, then m_1 Pareto dominates m_2 . Elements of the Pareto-front are models, that are not dominated by another model, that is, Pareto-optimal in terms of precision and memory cost.

4 New Scientific Results

Thesis group I: Achievements in Multi-directional Image Projections

Thesis 1.1

I have designed and implemented a method of mapping multi-directional projection vectors using fixed bin numbers regardless of the rotation angle. The memory cost of the result is independent of the image size; it is only affected by the rotation step number and the number of bins.

The computational complexity of the method based on trigonometric functions in the case of sequential processing is approximately the same as in the previous methods, determinable using the number of projection directions.

The length of projection slices is fixed even for different mapping directions; therefore, memory cost is known in advance. In the case of a low number of bins, the method can be interpreted as compression, providing a mapping with a lower memory cost than the Radon transformation. A generated output of the method is shown on Figure 4.1.

Publications pertaining to thesis: [1], [2], [3].

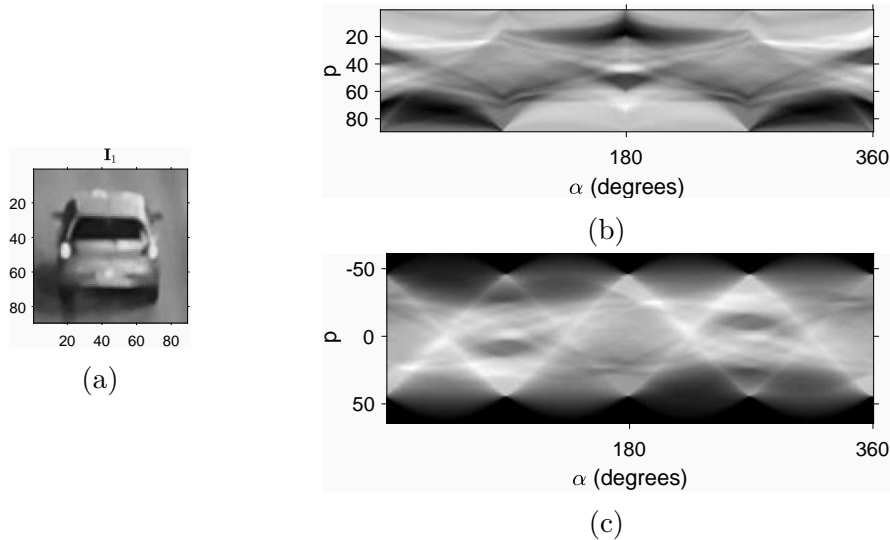


Figure 4.1: A sample output of the projection method with fixed number of bins. Subfigure (a) gives the original image, subfigure (b) shows the projection map. For comparison, subfigure (c) shows the sinogram of the Radon transform.

Thesis 1.2

I have designed and implemented the data-parallel version of the multi-directional image projection algorithm for graphical processors, which allows acceleration proportional to the number of execution units.

During the design of the solution, I took the specific memory architecture of the GPU devices into account, and sought to minimize the memory transfers between the host and the device. In order to avoid a race condition between common variables, I decided to redesign the algorithm, so it is not necessary to use synchronization to ensure correctness. When accumulating the results, atomic operations were used to secure correct behavior.

The algorithm is based on the division of the input matrix into logical blocks, which are stored in the *on-chip* memory of the multiprocessor of the GPU. The blocks are processed simultaneously, and the produced results are accumulated with the offset calculated from the original position of the blocks.

I compared the runtime of the parallel implementation with the sequential version and with the GPU-accelerated Radon transformation of *MATLAB*, where, in both cases, the mapping to projections was done in a shorter time if the number of bins was smaller than the image size.

Publications pertaining to thesis: [2], [3].

Thesis 1.3

I evaluated the effectiveness of the fixed vector length multi-directional image projection method for object matching, comparing the results with similar projection-based, lower-dimensional image signatures, and concluded that matching accuracy increased significantly.

I compared the defined method with two- and four-dimensional projection signature-based matching methods, as well as the Radon transform. Results show that

the distance between the true and false pairs of similarity scores for methods with a fixed number of bins is greater than in case of methods based on 2D or 4D projection signatures. The method was also compared with the higher-dimensional projection signature Radon transformation with the same result.

As a result of fixed resolution, memory cost is independent of the size of the image; therefore, it is unnecessary to use a sliding window to find the best fit, thus reducing potential errors.

Experiments with different resolution projections also show that the performance of the classification does not decrease even when using a low number of bins, resulting in an efficient solution in terms of memory cost.

Publications pertaining to thesis: [3].

Thesis group II: Application of Image Projections as Preprocessing in Siamese Convolutional Neural Networks

Thesis 2.1

I have developed a method based on backtracking search that provides all of the suitable convolutional neural network architectures at a given input, layer number, and memory cost.

After examining modern convolutional design patterns, I defined the requirements for the layers, which is the basis of the optimization to be performed. In the solution, the size of the activation maps is determined layer by layer, followed by the search for the next layers. Backtracking search eliminates the need to review all options, as layer combinations that do not lead to a valid solution are skipped.

The algorithm first determines the kernel and window sizes of the convolutional and pooling layer pairs, and then estimates the expected memory cost of the model based on the number of parameters. By defining an upper limit for memory cost, architectures optimized for target hardware can be generated.

After estimating the memory occupancy of the model, the batch sizes used in batch training can be optimized. With a larger batch size, the runtime of training is shorter.

Publications pertaining to thesis: [4].

Thesis 2.2

I designed and implemented a Master/Worker model for the analysis of Siamese convolutional neural network architectures in a distributed environment, with scheduling based on the longest processing times. In practical measurements, the parallel efficiency of the processing of the generated neural network architectures was 99.87%.

Distributed training took place in a cluster of computers equipped with graphical accelerators. I created the parallel solution based on the *Master/Worker* design pattern.

To optimize load balancing, I developed scheduling using LPT⁵ heuristics based on complexity estimation. I determined the training complexity of the models based on the number of trainable parameters and by the size of the batch.

Measurements have verified that the scheduling based on complexity estimation is effective, with the speedup of the entire process being almost equal to the number of workstations involved. The relationship between calculated complexities and measured processing times is strong, with a correlation coefficient of 0.749.

Based on the measured processing times, simulations were performed to investigate the efficiency of scheduling, and I found that automatic load balancing of the parallel method results in efficiency of over 90%, even with random task assignment; however, the same property using the presented scheduling is 99.87%.

Publications pertaining to thesis: [5].

Thesis 2.3

I analyzed each of the multidirectional image projection methods, using them as a preprocessor for input data, to determine the effect on the performance of Siamese convolutional networks. Based on the results, I concluded that the method based on a fixed number of bins is Pareto optimal in terms of efficiency and memory requirement compared to the raw image methods considered as a reference.

During the experiment the Radon, Trace and the transformations I defined were compared with different parameterizations. Based on the dimensions of the defined mappings, Siamese convolutional architectures were generated and processed in a distributed environment.

The object matching accuracy of the trained models was determined using *one-shot* classification tests, where the number of objects compared to the reference was $2 \leq N \leq 10$.

After evaluating the results in terms of accuracy and memory occupancy, I concluded that the method based on fix bin number is a Pareto optimal choice contrast to the methods based on image input. The results of the comparison is visualized on Figure 4.2.

Publications pertaining to thesis: [6].

5 Practical Applicability of the Results

Object identification or re-identification ⁶ is an important field, not restricted to tracking and matching vehicles as discussed in the dissertation. Common solutions exist for access control and outdoor camera systems to identify people.

The greatest advantage of the developed method is that it fits perfectly into today's IoT⁷ smart camera systems. Beyond capturing the image, the camera also processes it: moving objects are detected and segmentation and classification is done. As the last step, the projection signature of the object can also be computed and transferred over the network connection to another unit.

⁵Longest Processing Times

⁶object reidentification

⁷Internet of Things

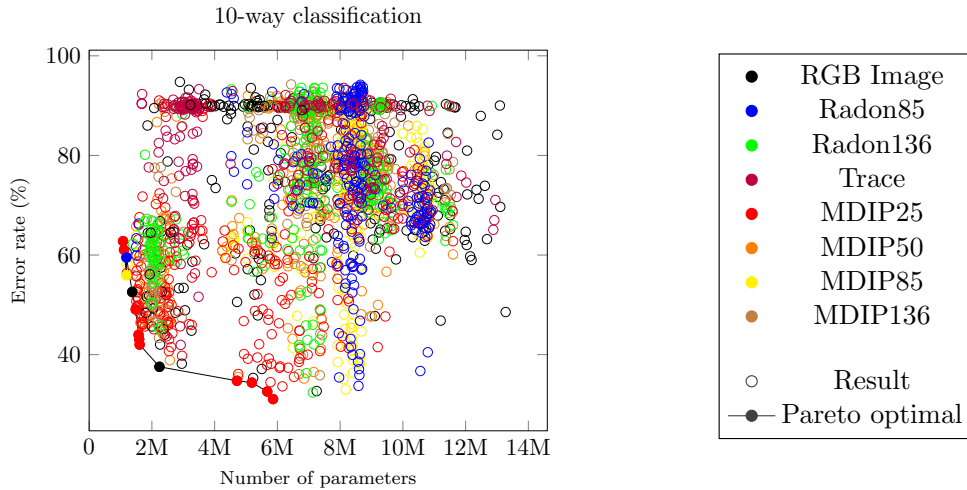


Figure 4.2: Model performances in terms of parameter number and the accuracy of the 10-way one-shot classification measured on the validation dataset. Each model was tested with 10000 validation measurements. The Pareto optimal models – which are not dominated by any other element – are on the lower left corner as Pareto-frontier.

The advantage of a fixed number of bins is that the memory cost is independent of the input image size, so communication costs can be well determined. If a low-resolution projection map is produced, the method behaves as compression.

The parallel solution provides fast response times even in case of large images.

The method works well for *one-shot* classification problems: it is also able to detect objects that were not encountered earlier during the training phase.

The method of generating neural network architectures developed in the study of machine learning-based method is capable of generating a structure of any number of convolutional and pooling layer pairs for any input matrix within a predefined memory limit. Therefore, the solution is applicable in other cases where more memory-critical neural architecture may be required.

Bibliography

- Bay, Herbert et al. (2008). “Speeded-up robust features (SURF)”. In: *Computer vision and image understanding* 110.3, pp. 346–359.
- Bengio, Yoshua (2012a). “Deep learning of representations for unsupervised and transfer learning”. In: *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, pp. 17–36.
- (2012b). “Practical Recommendations for Gradient-Based Training of Deep Architectures”. In: *Neural networks: Tricks of the trade*. Springer, pp. 437–478.
- Betke, Margrit, Esin Haritaoglu, and Larry S Davis (2000). “Real-time multiple vehicle detection and tracking from a moving vehicle”. In: *Machine vision and applications* 12.2, pp. 69–83.
- Bromley, Jane et al. (1994). “Signature verification using a " siamese" time delay neural network”. In: *Advances in neural information processing systems*, pp. 737–744.
- Brunelli, R. (2009). *Template Matching Techniques in Computer Vision: Theory and Practice*. Wiley.
- Choe, Tae Eun, Mun Wai Lee, and Niels Haering (2010). “Traffic Analysis with Low Frame Rate Camera Networks”. In: *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*. IEEE, pp. 9–16. ISBN: 978-1-4244-7030-3. DOI: 10.1109/CVPRW.2010.5543801.
- Chopra, Sumit, Raia Hadsell, and Yann LeCun (2005). “Learning a similarity metric discriminatively, with application to face verification”. In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. Vol. 1. IEEE, pp. 539–546.
- Deans, Stanley R. (1983). *The Radon Transform and Some of Its Applications*. New York: John Wiley and Sons.
- Ginkel, M. van, C.L. Luengo Hendriks, and L.J. van Vliet (2004). *A short introduction to the Radon and Hough transforms and how they relate to each other*. Tech. rep. QI-01-2004.
- Goodfellow, Ian et al. (2016). *Deep learning*. Vol. 1. MIT press Cambridge.
- Guo, Yanlin, Steve Hsu, et al. (2007). “Robust Object Matching for Persistent Tracking with Heterogeneous Features”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29.5, pp. 824–839. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2007.1052.
- Guo, Yanlin, Cen Rao, et al. (2008). “Matching vehicles under large pose transformations using approximate 3D models and piecewise MRF model”. In: *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, p. 8.
- Hou, Tingbo, Sen Wang, and Hong Qin (2009). “Vehicle Matching and Recognition under Large Variations of Pose and Illumination”. In: *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*. IEEE, pp. 24–29. ISBN: 978-1-4244-3994-2. DOI: 10.1109/CVPRW.2009.5204071.
- Hough, Paul VC (Dec. 1962). *Method and means for recognizing complex patterns*. US Patent 3,069,654.

- Jelača, Vedran et al. (2013). “Vehicle matching in smart camera networks using image projection profiles at multiple instances”. In: *Image and Vision Computing* 31, pp. 673–685.
- Kadyrov, Alexander and Maria Petrou (2001). “The trace transform and its applications”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23.8, pp. 811–828.
- Karpathy, Andrej, FF Li, and J Johnson (2017). “CS231n: Convolutional Neural Networks for Visual Recognition, 2016”. In: URL <http://cs231n.github.io>.
- Koch, Gregory, Richard Zemel, and Ruslan Salakhutdinov (2015). “Siamese Neural Networks for One-shot Image Recognition”. In: *ICML 2015 Deep Learning Workshop*.
- Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton (2012). “ImageNet Classification with Deep Convolutional Neural Networks”. In: *Proceedings of the 25th International Conference on Neural Information Processing Systems. NIPS’12*. Lake Tahoe, Nevada: Curran Associates Inc., pp. 1097–1105.
- LeCun, Yann A et al. (2012). “Efficient BackProp”. In: *Neural networks: Tricks of the trade*. Springer, pp. 9–48.
- LeCun, Yann, Yoshua Bengio, et al. (1995). “Convolutional networks for images, speech, and time series”. In: *The handbook of brain theory and neural networks* 3361.10.
- LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton (2015). “Deep learning”. In: *Nature* 521.7553, p. 436.
- LeCun, Yann, Bernhard Boser, et al. (1989). “Backpropagation applied to handwritten zip code recognition”. In: *Neural computation* 1.4, pp. 541–551.
- Lee, Seungkyu, Yanxi Liu, and Robert Collins (2007). “Shape variation-based frieze pattern for robust gait recognition”. English (US). In: *2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR’07*. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. ISBN: 1424411807. DOI: 10.1109/CVPR.2007.383138.
- Lindeberg, Tony (1998). “Feature detection with automatic scale selection”. In: *International journal of computer vision* 30.2, pp. 79–116.
- Liu, Yanxi, Robert Collins, and Yanghai Tsin (2002). “Gait sequence analysis using frieze patterns”. In: *European Conference on Computer Vision*. Springer, pp. 657–671.
- Lowe, David G et al. (1999). “Object recognition from local scale-invariant features”. In: *Proceedings of the International Conference on Computer Vision*. 2, pp. 1150–1157.
- Oren, Michael et al. (1997). “Pedestrian detection using wavelet templates”. In: pp. 193–199.
- Oriol Vinyals and Charles Blundell and Timothy P. Lillicrap and Koray Kavukcuoglu and Daan Wierstra (2016). “Matching Networks for One Shot Learning”. In: ed. by D. D. Lee et al., pp. 3630–3638.
- Papageorgiou, Constantine P, Michael Oren, and Tomaso Poggio (1998). “A general framework for object detection”. In: *Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271)*. IEEE, pp. 555–562.
- Radon, Johann (1917). “Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten”. In: *Berichte über die Verhandlungen*

- der Königlich-Sächsischen Akademie der Wissenschaften zu Leipzig, Mathematisch-Physische Klasse*, pp. 262–277.
- Raina, Rajat, Anand Madhavan, and Andrew Y. Ng (2009). “Large-scale Deep Unsupervised Learning Using Graphics Processors”. In: *Proceedings of the 26th Annual International Conference on Machine Learning*. ICML ’09. Montreal, Quebec, Canada: ACM, pp. 873–880. ISBN: 978-1-60558-516-1. DOI: 10.1145/1553374.1553486.
- Rios-Cabrera, Reyes, Tinne Tuytelaars, and Luc Van Gool (2012). “Efficient Multi-camera Vehicle Detection, Tracking, and Identification in a Tunnel Surveillance Application”. In: *Comput. Vis. Image Underst.* 116.6, pp. 742–753. ISSN: 1077-3142. DOI: 10.1016/j.cviu.2012.02.006.
- Sanchez, Angel et al. (2011). “Video-Based Distance Traffic Analysis: Application to Vehicle Tracking and Counting”. In: *Computing in Science and Engg.* 13.3, pp. 38–45. ISSN: 1521-9615. DOI: 10.1109/MCSE.2010.143.
- Schmid, Cordelia, Roger Mohr, and Christian Bauckhage (2000). “Evaluation of interest point detectors”. In: *International Journal of computer vision* 37.2, pp. 151–172.
- Schmidhuber, Jürgen (2015). “Deep learning in neural networks: An overview”. In: *Neural networks* 61, pp. 85–117.
- Schroff, Florian, Dmitry Kalenichenko, and James Philbin (2015). “FaceNet: A Unified Embedding for Face Recognition and Clustering”. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2015* abs/1503.03832.
- Shan, Ying, Harpreet S. Sawhney, and Rakesh Kumar (2005). “Vehicle Identification between Non-Overlapping Cameras without Direct Feature Matching”. In: *10th IEEE International Conference on Computer Vision (ICCV’05)* 1, pp. 378–385.
- Shi, Jianbo and Carlo Tomasi (1993). *Good features to track*. Tech. rep. Cornell University.
- Szegedy, Christian et al. (2015). “Going Deeper with Convolutions”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9.
- Szeliski, Richard (2010). *Computer Vision: Algorithms and Applications*. 1st. New York, NY, USA: Springer-Verlag New York, Inc. ISBN: 1848829345, 9781848829343.
- Taigman, Yaniv and Yang, Ming and Ranzato, Marc’Aurelio and Wolf, Lior (2014). “DeepFace: Closing the Gap to Human-Level Performance in Face Verification”. In: *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition*. CVPR ’14. Washington, DC, USA: IEEE Computer Society, pp. 1701–1708. ISBN: 978-1-4799-5118-5. DOI: 10.1109/CVPR.2014.220.
- Viola, Paul and Michael Jones (2001a). “Rapid Object Detection using a Boosted Cascade of Simple Features”. In: *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. Vol. 1. IEEE, pp. 511–518.
- (2001b). “Robust real-time object detection”. In: *International journal of computer vision* 4.34-47, p. 4.
- Yuan, Quan and Stan Sclaroff (2009). “Is a detector only good for detection?” In: *2009 IEEE 12th International Conference on Computer Vision*. IEEE, pp. 1066–1073.

Own Publications Pertaining to Theses

- [1] Kertész, Gábor, Sándor Szénási, and Zoltán Vámosy (2017). “Application and properties of the Radon transform for object image matching”. In: *Proceedings of SAMI 2017. IEEE 14th International Symposium on Applied Machine Intelligence and Informatics* (Herlany, Slovakia, Jan. 26–28, 2017). IEEE, pp. 353–358.
- [2] — (2016). “A Novel Method for Robust Multi-Directional Image Projection Computation”. In: *Proceedings of INES 2016. 20th IEEE International Conference on Intelligent Engineering Systems* (Budapest, Hungary, June 30–July 2, 2016). IEEE, pp. 239–243.
- [3] — (2018). “Multi-Directional Image Projections with Fixed Resolution for Object Matching”. In: *Acta Polytechnica Hungarica* 15.2, pp. 211–229.
- [4] — (2019a). “A novel method for Convolutional Neural Architecture Generation with memory limitation”. In: *Proceedings of SAMI2019. IEEE 17th World Symposium on Applied Machine Intelligence and Informatics* (Herlany, Slovakia, Jan. 24–26, 2019). IEEE, pp. 229–234.
- [5] — (2019b). “Distributed training and evaluation of projection-based descriptors in Siamese Neural Networks”. In: *Proceedings of the Sixth International Conference on Parallel, Distributed, GPU and Cloud Computing for Engineering*, Paper 25, 1–12. DOI: 10.4203/ccp.112.25.
- [6] — (2019c). “Multi-Directional Projection Transformations for Machine Learning based Object Matching”. In: *SACI 2019 : IEEE 13th International Symposium on Applied Computational Intelligence and Informatics*, pp. 269–274.

Other Publications

- [7] Kertész, Gábor, Sándor Szénási, and Zoltán Vámosy (2015a). “Performance Measurement of a General Multi-Scale Template Matching Method”. In: *Proceedings of INES 2015. 19th IEEE International Conference on Intelligent Engineering Systems* (Bratislava, Slovakia, Sept. 3–5, 2015). IEEE, pp. 153–158.
- [8] — (2015b). “Parallelization Methods of the Template Matching Method on Graphics Accelerators”. In: *Proceedings of CINTI 2015. 16th IEEE International Symposium on Computational Intelligence and Informatics* (Budapest, Hungary, Nov. 19–21, 2015). IEEE, pp. 161–164.
- [9] Kertész, Gábor, Szabolcs Sergyán’, et al. (2016). “Implementation of Object Recognition based on Image Projection Signatures using Matlab”. In: *Proceedings of CINTI 2016. 17th IEEE International Symposium on Computational Intelligence and Informatics* (Budapest, Hungary, Nov. 17–19, 2016). IEEE, pp. 123–128.
- [10] Kertész, Gábor, Sándor Szénási, and Zoltán Vámosy (2018b). “Vehicle Image Matching Using Siamese Neural Networks with Multi-Directional Image Projections”. In: *12th IEEE International Symposium on Applied Computational Intelligence and Informatics, SACI 2018, Timisoara, Romania, May 17-19, 2018*, pp. 491–496.
- [11] Kertesz, Gabor and Eva Hajnal (2013). “Irisz Project: A Web Application for the Introduction of University Students to the Labor Market”. In: *International Symposium on Applied Informatics and Related Areas : AIS 2013 Szekesfehervar, Magyarország : Óbudai Egyetem, (2013)*, pp. 125–129.
- [12] — (2014). “Special Issues in the Development of a Large User Based Web Application”. In: *Proceedings of the IEEE 12th International Symposium on Applied Machine Intelligence and Informatics (SAMI 2014) Budapest, Magyarország : IEEE Hungary Section, (2014)*, pp. 141–145.
- [13] Kertész, Gábor and Zoltán Vámosy (2015). “Current challenges in multi-view computer vision”. In: *10th IEEE Jubilee International Symposium on Applied Computational Intelligence and Informatics, SACI 2015, Timisoara, Romania, May 21-23, 2015*, pp. 237–241.
- [14] — (2016). “A Brief Review of Recent Advances in Multi-View Computer Vision”. In: *Scientific Bulletin of Politechnica University of Timisoara - Transactions on Automatic Control and Computer Science* 61(75), pp. 73–78. ISSN: 1224-600X.
- [15] Kertesz, Gabor, Daniel Kiss, et al. (2016). “Multiprocessing of an individual-cell based model for parameter testing”. In: *11th IEEE International Symposium on Applied Computational Intelligence and Informatics, SACI 2016, Timisoara, Romania, May 12-14, 2016*, pp. 491–496.
- [16] Szenasi, Sandor et al. (2018a). “Comparison of Road Accident Black Spot Searching Methods”. In: *IEEE 18th International Symposium on Computational Intelligence and Informatics (CINTI 2018) Budapest, Magyarország : IEEE Hungary Section, (2018)*, pp. 247–250.

- [17] Szenasi, Sandor et al. (2018b). “Road Accident Black Spot Localisation using Morphological Image Processing Methods on Heatmap”. In: *IEEE 18th International Symposium on Computational Intelligence and Informatics (CINTI 2018) Budapest, Magyarország : IEEE Hungary Section, (2018)*, pp. 251–256.
- [18] Danko, Bence and Gabor Kertesz (2018a). “Recognition of the Hungarian Fingerspelling Alphabet using Convolutional Neural Network based on Depth Data”. In: *IEEE 18th International Symposium on Computational Intelligence and Informatics (CINTI 2018) Budapest, Magyarország : IEEE Hungary Section, (2018)*, pp. 41–46.
- [19] — (2018b). “Recognition of the Hungarian fingerspelling alphabet using Recurrent Neural Network”. In: *SAMI 2019 : IEEE 17th World Symposium on Applied Machine Intelligence and Informatics*, pp. 251–256.

