

Computer Vision and Remote Sensing – Lessons Learned

WOLFGANG FÖRSTNER, Bonn

ABSTRACT

Photogrammetry has significantly been influenced by its two neighbouring fields, namely Computer Vision and Remote Sensing. Today, Photogrammetry has become a part of Remote Sensing. The paper reflects its growing relations with Computer Vision, based on a more than 25 years experience of the author with the fascinating field between cognitive, natural and engineering science, which stimulated his own research and transferred him into a wanderer between two worlds.

“Two roads diverged in a wood, and I –
I took the one less traveled by, and that has made all the difference.”
Robert Frost

1. INTRODUCTION

My first contact with a scientific community not being of geodetic nature was in 1983. Franz Leberl had organized the Specialist Workshop on *Pattern Recognition in Photogrammetry* in Graz (Leberl 1984) on behalf of the International Society of Photogrammetry and Remote Sensing by its Working Group III/5 *Mathematical Pattern Recognition and Image Analysis*. Well known photogrammetrists presented papers, such as Ed Mikhail or Ian Dowman. But also researches from a – that time – foreign world, such as Harlyn Baker or Robert Haralick, gave a flavour of techniques, not known in the core photogrammetric community, though addressing the same topics, specifically automatic surface reconstruction. The stimulating discussions with people from Pattern Recognition were the starting point for regular travels into and through a fascinating world around the own origin, geodesy and photogrammetry, which is still ongoing, full in the spirit of the workshop: “ ... encourage future participation in pattern recognition ...”. The journey changed my views on my origins and made them richer.

The journey required to learn a new language. Already in 1983 notions as “Expert systems” or “Image Understanding” used to describe basic concepts in pattern recognition were not “part of photogrammetry’s jargon”. It took me a few years to understand the language of the other discipline and about six years to roughly speak the language. In the meantime both fields evolved: all those, who analyzed images came together in workshops on *Computer Vision*, however, also inspired by the upcoming field of *Artificial Intelligence*, and the availability of satellite imagery made Photogrammetry a part of *Remote Sensing*. As a consequence, the number of languages and cultures increased, making communication even more difficult. Such a multilingual and multicultural situation uncovers characters: Who learns the other language? How do people react on non-native speakers? How do notions change by regularly talking across language borders? How are techniques transferred, modified and adapted? Does there evolve a common understanding?

In the following I want to share experiences from that travel. I first want to sketch my view on the developments of our own and of our neighbor’s field, illuminating the boundaries by some personal

experiences. The cautious, sometimes hesitating steps of photogrammetrists outside their own field resulted in partially contradicting views which need to be made aware in order to smooth the way for a further increase of the mutual interest, and understanding between our highly active field and the country around us which hopefully will lead to intensive cooperations in the future.

2. DEVELOPMENTS

2.1. Photogrammetry and Remote Sensing

While Photogrammetry, as an engineering discipline, can be tracked back to the 19th century, the development of Remote Sensing started with the first satellites in the 70's.

Though some photogrammetric institutes grasped the at that time highly innovative possibilities for earth observation from satellites, especially to solve the urgent mapping problem, remote sensing developed nearly independently. For the first time geosciences had access to supra-regional observation platforms. Digital images allowed evaluating the information using automatic classification methods and linking these data with various models for geo-processes.

Only in the 80's, after proving knowing the accuracy potential of image matching procedures, photogrammetry started to apply image processing. Today all classical photogrammetric processes are automated, starting with the last step in the production chain orthophoto production, then generation of digital elevation and surface models and finally aerial triangulation, today even including the automatic orientation in close range applications.

Today there is no need to distinguish between the two fields, as due to the overlap in spectral and spatial resolution of remote sensing and photogrammetric sensors techniques melt together. Only representatives of different applications show different culture in exploiting the potential.

This view, however, does not fully reflect reality: there are still mutual preconceptions. As an example: some photogrammetrists still hesitate to accept the interpretation of satellite images with pixel sizes larger than 30 m as a topic of their field, and only accept scientific results if the precision inherent in the images is exploited. Some remote sensors still hesitate to accept the photogrammetrists' questions about the consistency of geometric evaluation of satellite images and only accept scientific results, if a geoscientific application is addressed.

I take this situation as an indication for the end of a transition towards a fruitful mutual appreciation, a vision which those, who decided to name societies according to both fields, Photogrammetry and Remote Sensing, had long before. The relation to the fields of Pattern Recognition and Computer Vision appears to be in a much earlier stage.

2.2. Pattern Recognition and Computer Vision

2.2.1 The eighties. The International Association for Pattern Recognition was founded in 1978 two years after the foundation of the German for Pattern Recognition (DAGM). At the same time (1979), Ulf Grenander at Brown University published a book on 'Pattern Theory', laying the theoretical ground for many later developments in image understanding, including Markov random fields and grammars, however, incomprehensible due to its cryptic notation. Only four years later, in 1982, the first Workshop on Computer Vision initiated a series of successful events. The topic 'Representation and Control' already then indicated the basic problems when making machines see:

“How is reality represented in a computer?” and “What roles have embedded vision systems for learning machines?” The military funding of vision research in the USA initiated a long series of DARPA Workshops on Image Understanding. The Computer Vision in the 80’s was driven by the idea to build intelligent seeing machines, on one side dramatically underestimating the complexity of the task, on the other hand motivating a discussion on the theoretical foundations of the discipline. David Marr, at MIT, having studied mathematics and neuro-informatics, certainly was most influencing (1982): The brain is a computer analyzing the sensor data provided on the retina and building an internal map of the outside world. His “primal sketch” stimulated work on image feature extraction, especially feature based 3D reconstruction methods. It took two decades until this view was overcome, and vision – at least by parts of the community – was understood as part of a system embedded and living in the real world, continuously learning the role of sensoric stimuli. This was in contrast to the view, driven by the field of Artificial Intelligence, which culminated in the Japanese Fifth Generation Computer Systems project, where all intellectual problems were supposed to be solvable based on logic and in parallel stimulated various image analysis systems, including quite some on interpreting aerial images (cf. Herman et al. 1983, McKeown 1988, Matsuyama and Hwang 1990).

2.2.2 The nineties. The vision of the 80’s failed. The 90’s went back to the roots – satisfyingly for photogrammetrists, seen from now – to projective geometry. The bi-annual International Conference on Computer Vision, the first being in 1989, and its European counterpart with more than a third of the papers each year covered all aspects of geometric reasoning and produced photogrammetry at its best – seen from inside. Three textbooks (Faugeras et al. 1993, Hartley and Zisserman 2002, Faugeras 2004) documented the achievements, now are a must when studying Computer Vision – and Photogrammetry. The scientific interest in geometry since then has shifted toward general especially omnidirectional cameras, which went along with innovative hardware developments presented at the international conferences. Parallel to these developments, mathematicians and theoretical physicists interested in Computer Vision became influential. Already in 1984 the key paper by S. and D. Geman, scholars of U. Grenander, on Markov Random Fields picked up the models from statistical mechanics and integrated them in a Bayesian framework for image restoration, a methodology which up to now is in the core of all image interpretation research. In contrast to these methods, which refer to graphs as basic representation, in 1988 variational methods, where instead of parameter vectors functions are to be estimated, were proposed by Osher and Sethian to solve inverse problems, Mumford and Shah (1989) set the theoretical basis for image segmentation. Today all basic Computer Vision algorithms, such as segmentation, optical flow or stereo reconstruction are successfully attacked using these strong tools.

2.2.3 The new century. D. Mumford, owner of the fields medal 1974, the highest award in mathematics, opened the discussion on the mathematical foundations at the beginning of the millennium with a strong vision (2000): past mathematics was deterministic, future mathematics will be stochastic: “stochastic models and statistical reasoning are more relevant i) to the world, ii) to science and many parts of mathematics and iii) particularly to understanding the computations in our own minds, than exact models and logical reasoning.” Observing the trend in Computer Vision Conferences during the last decade confirms this: nearly half of the papers deal with statistically founded methods. Moreover, the topic is object detection, classification, image interpretation or behaviour analysis from videos. As an example, one great challenge is to identify the existence of one of 256 objects in an image taken from a nearly arbitrary viewpoint and with arbitrary background. Application is content based image retrieval: finding a non-annotated image based on its visual content only.

The broadness of the field is obvious, not directly its relevance for photogrammetry.

2.3. Some experiences on the bridge

Let me give some personal views I had during this time.

1983: I already mentioned the start in Graz's workshop on 'Pattern Recognition in Photogrammetry.' I was puzzled by the new notions I did not understand. I remember Ian Dowmans idea of saliency when finding key points in matching, wondering how one can link such a vague notion with a mathematical concept in a unique manner.

1984: The discussion with Ed Mikhail resulted in a three months visit of Purdue, a university town in the middle of nowhere, a library which had office hours until midnight and a windowless working room, where one forgot the time. I had the chance to read and visit a one week course on Pattern recognition and Image processing. Fukunaga gave an excellent lecture on statistical pattern recognition, cf. his still highly recommended book (2003), King Sun Fu's lecture on structural pattern recognition appeared to be too far off from any application, but contained texture models and especially stochastic grammars, one of the key models for today's methods for interpreting images of man made scenes. That time I decided not to work on image correlation and on adjustment theory any more, unless I urgently need it, in order to have time for learning new and more and at least equally important techniques. That year also the ISPRS WG III/4 test on image matching started: Marsha Jo Hanna, from Stanford Research Institute won the contest: it was disappointing, as her method was not based on a theory, but was the result of clever engineering.

1985: The participation at the Third Workshop on Computer Vision 'Representation and Control' being the first Computer Vision event for me, probably had most influence on my thinking. Avi Kak, coauthor of the Image processing book with Rosenfeld, invited me, to share the room to reduce my travel expenses (I had no paper) – openness reached me. In a discussion I naively asked what 'idgekai' means¹, resulting in a moment of silence – as if in this audience somebody would ask what 'aiaspearas' means. Times were turbulent. Bob Haralick's talk on 'Computer Vision theory: the lack thereof' and Keith Price's presentation 'I have seen your demo - so what?' indicated the field not to be so mature, as some of its representatives would have liked to see it.

1988: Visiting the the Artificial Intelligence Center at Stanford Research Institute opened my eyes: They *did* Digital Photogrametry, though this term was coined later. Fua and Hanson (1988) built a system for automatically detecting the ground plan of complex buildings, a first instance of a grammar used for aerial image interpretation. They used so-called snakes to extract roads, of course used stereo matching for deriving DEM's, but most impressive had a tool to interactively extract the 3D-structure of buildings, all integrated in a digital system with stereo viewing.

1989: E. Gülch and me submitted a paper on detecting points to the first ICCV. We received three slips of paper with comments on it, the first (negative) experience with a double blind review process: the language of our paper could not transport the idea. We learned: the paper with Monika Sester on building location, submitted to the DAGM meeting, was written densely in the right language and awarded. Since then I joined worked with and for the German Association for Pattern Recognition, a second home.

¹ IJCAI: International Joint Conference for Artificial Intelligence, the leading conference in AI.

1990: I started to teach Photogrammetry. The intention was to allow students to communicate with their Computer Science friends on scientific meetings and to motivate these students to take Photogrammetry as a second subject. I took the maths of the Computer Vision and applied our thinking: The collinearity equation is easier to handle in the form $x' = PX$, than in the long form with ratios of polynomials, but still allows to work with cameras where the interior orientation is known. The discussions with my photogrammetric colleagues appears not really closed at this point. I waited for the direct solution of the relative orientation, told my students they should come if they had one, as this would be a breakthrough in image orientation – it came ten years later with the neat solution of D. Nister (2003). The lectures were attended by four to eight computer science students each year, some worked on their PhD in Photogrammetry, since this year we offer a module for their bachelor.

1993-1999: The DFG-bundle project ‘Semantic Modeling’ gave the chance to start to cooperate with colleagues from Computer Vision and Pattern Recognition. With the Ascona workshops in 1995 and 1997 the topic became widespread. The interdisciplinary work appeared to be successful: young researchers in the US were asked to observe what happened in Europe.

1996: Becoming a member of the Program committee of the European Conference of Computer Vision gave a close insight into the flavour of the double blind review process of half a thousand papers: I could experience how to make fair decisions on paper selections in a one day concentrated meeting – at the same time observing the possibility to evaluate the reviewers (later, since the ICCV 2005 the best reviewer was given an award). At one of the coffee breaks A. Zisserman asked me whether I knew something about constraints between three images. I did not really understand the question – they wanted to know whether the generalization of the epipolar constraint for three images, specifically the trifocal tensor, was known to photogrammetrists. When preparing a paper for the next ECCV I remembered the PhD of Ed Mikhail on the image triplet, and found that he published a constraint on three images in 1962 and 1963! It was hidden and not in the awareness.

2002-2005: At a review meeting for EU-projects in 2001 the idea of a European research network was born: The European network for cognitive computer vision systems – ECVision – was inaugurated. The semiannual meetings of a highly interdisciplinary group finally lead to a 15 years perspective, compressed into a roadmap on research in Cognitive Computer Vision. It lays open the complementary views on Computer Vision, either starting learning to see from scratch following an evolutionary approach or assuming structural preconditions which allow logical/statistical reasoning. It also was the motivation to apply for a 3-years EU project on ‘E-Training for Interpreting Images of Man-Made Scenes’ (eTRIMS) to start to understand the flavour of statistical learning theory for the use in image interpretation: one of eight PhD-students is a photogrammetrist.

What lessons did I learn?

3. SOME LESSONS LEARNED

3.1. The size of research communities

The size of a research community matters. The peaks of a sample are larger for larger samples. The number of innovative ideas, the scientific level, the likelihood to find a solution to a trendy problem increases with the size of a research community. The virtuosity of handling mathematics, physics and geometry, the clarity in writing, the persuasiveness of presentations is stimulating at the same time making modest. It increases the scale of evaluation in both directions, always finding someone to learn from, but also better identifying the own strengths.

As a consequence there is need for the smaller group to go to the larger, not vice versa. We as Photogrammetrists need to go to their conferences, submit papers, show our profile, self-critically adapt our language. Quite some researchers in Computer Vision appreciate Photogrammetric research and come to our conferences. But this is the exception, trying to increase the mutual engagement during my time as ISPRS Commission III president was only partially successful: More photogrammetric researchers get involved in the neighbouring conferences, the number of people from Computer Vision visiting our conferences still is small.

As soon as we have something to tell, they realize: the excellent overview on the bundle adjustment by Triggs et al. (2000) could not be written clearer and covers all relevant aspects, also demonstrating that the prejudice Computer Vision people reinvent what Photogrammetry knows since a long time is not true. And if a computer scientist asked, where to find the essentials of photogrammetry, I could not give an answer – until 2004, when the fifth edition of the ASPRS Manual of Photogrammetry appeared. Our experience was buried in unreviewed proceedings and the journal papers were mostly written by geodesist for geodesists.

3.2. The degree of interdisciplinarity

Computer Vision has at least two aspects. It is an engineering discipline aiming a working solutions and it is a natural science discipline aiming at understanding the human visual system. As a consequence there is a fluctation in arguments. A strong one is the biology inspired motivation for certain structures of complete systems or of special low level processes. As an example J. Tsotsos (1988) analysed the complexity of immediate vision: how can the human visual system with process with a processing speed of one step per milli-second come to a decision in a tenth of a second? On the other hand the well known Lowe operator (2004) was designed after long years studies of the human visual system, not aiming at mimicing the processing on the neural level. Innovations happen when crossing borders of disciplines. We can refresh our own research potential by joining interdisciplinary groups, where each participant appreciates the capacity of the others and accepts that one discipline cannot solve the complex problem of image interpretation.

3.3. Commons and Non-Commons

The technical problems of Photogrammetry and Remote Sensing and of Vomputer Vision and Pattern Recognition overlap to a large content, as layed open. When I started to read the first volumes of the Transactions on Pattern Analysis and Machine Intelligence in the 80's my immediate reaction was: At least half of the papers are relevant, I would like to understand them. This has not changed since. We can exploit the large potential of methods to our advantage – provided we work hard in improving our mathematical background, this is a prerequisite. Therefore any reduction in mathematical education is counterproductive: it is easier to learn the specific

boundary conditions and specialities of the various – and ever changing – applications, than to learn the mathematics, especially statistics: adjustment theory, only covering a tenth of the necessary statistical tools, is by far not sufficient.

The main difference between our field and Computer Vision is the motivation and the research goals: They aim at establishing strong methods, afterwards searching for an application, relying there are some, actually there are more applications than can be addressed. We have an application problem: solving the problem of automatic mapping from images. We not only know the relevance but have the infrastructure both in science as in practice to immediately transfer knowledge to the users. The Photogrammetric Week is one of its most prominent examples.

3.4. Openness

Research essentially is communicating verified results which are relevant to others. This explains the current trend in evaluating the quality of research via publications – without discussing the sometimes doubtful measures used. The self-awareness of photogrammetric research appeared highest – at least to me – in the area where aerial triangulation in all its forms was the hot topic. The reason for this awareness was the existence of common benchmarks, Oberschwaben and Appenweier, to name two. Benchmarks are common in Computer Vision, Pattern Recognition cannot do without it. Benchmarks are costly as ground truth is costly. Benchmarking requires openness, providing code (at least exe-files) for free. I would like to see this more in our field: The camera calibration test of the German Society for Photogrammetry (2009) is an example for such an effort – with the hope it is made international. Benchmarks for Pattern Recognition not seem to have reached Remote Sensing. However, the exchange of software has no tradition in Photogrammetry. Since I entered the Computer Vision field I was asked, whether there exists a low cost bundle adjustment program for scientific use. I could not tell. In 2004 a group of computer scientists made code of a general bundle adjustment freely available, cf. Lourakis and Argyros (2004). It exploits the numerical advantages of sparse matrices and allows entering any projection model.

4. CONTINUING THE DIALOGUE

The dialogue between our own field and the neighbouring disciplines is active and useful for both sides. As a small discipline we need to dive into the thinking of our colleagues: learning their structuring of the scientific field, learning their way of solving problems, learning their motivations for doing research and development. Experience with others always is colored by the own history. We need to continue the dialogue and share our experiences find a common view not only among us few but within a rich and fascinating group of people who work towards the same goal.

“To travel is to discover that everyone is wrong about other countries.”
Aldous Huxley

REFERENCES

- P. Auer et al. (2005): A Research Roadmap of Cognitive Vision, www.ecvision.org (under reconstruction at the time of writing, please ask the author for a copy).
- M. Cramer, H. Krauß, K. Jacobsen, M. Von Schönemark, N. Haala, V. Spreckels: Das DGPF-Projekt zur Evaluierung digitaler photogrammetrischer Kamerasysteme, DGPF-Tagungsband, 18/2009.
- O. Faugeras (1993): Three-Dimensional Computer Vision, The MIT-press, 1993.
- O. Faugeras, Q.-T. Luong, T. Papadopoulo (2004): The Geometry of Multiple Images, The MIT-Press, 2004.
- W. Förstner, E. Gülch (1987): A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centres of Circular Features, In: Proceedings of the ISPRS Conference on Fast Processing of Photogrammetric Data. Interlaken 1987, S. 281-305.
- W. Förstner, M. Sester (1989): Object Location Based on Uncertain Models, In: Burkhardt, K. H. / Höhne, B. Neumann, B. (Hg.): Mustererkennung 1989, 11. DAGM-Symposium. Hamburg 1989, S. 457-464.
- P. Fua and A. J. Hanson (1991): An Optimization Framework for Feature Extraction, Machine Vision and Applications, Vol. 4, Nr. 2, pp. 59-87, 1991.
- K. Fukunaga (2003): Statistical Pattern Recognition, Springer, 2003.
- S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE-PAMI, 6, 1984, pp. 721-741.
- U. Grenander (1979): General Pattern Theory: A mathematical Study of Regular Structures, Clarendon Press, 1979.
- R. Hartley, A. Zisserman (2002): Multiple View Geometry in Computer Vision, Cambridge University Press, 2002.
- M. Herman, T. Kanade, S. Kuroe (1983): The 3D MOSAIC Scene Understanding System. Int. Joint Conf. on Artificial Intelligence, 1983, pp. 1108-1112.
- F. Leberl (1984): Pattern Recognition in Photogrammetry, Special Issue, Photogrammetria, Vol. 39., No. 3, pp. 61 ff.
- M. I. A. Lourakis, A. A. Argyros (2004): The Design and Implementation of a Generic Sparse Bundle Adjustment Software Package Based on the Levenberg-Marquardt Algorithm, <http://www.ics.forth.gr/~lourakis/sba>, last visited 20. 7. 2009.
- David G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision, 60, 2 (2004), pp. 91-110.
- D. Marr (1982): Vision. A Computational Investigation into the Human Representation and Processing of Visual Information, W. H. Freeman and Company.

- T. Matsuyama, V. S.-S. Hwang (1990): SIGMA: A Knowledge-Based Aerial Image Understanding System, Plenum Press, 1990.
- D. M. McKeown Jr. (1988): Building Knowledge-Based Systems for Detecting Man-Made Structures From Remotely Sensed Imagery, *Phil. Trans. R. Soc. London. A(324)*, 1988, pp. 423-435.
- E. Mikhail (1962): Use of Triplets in Analytical Aerotriangulation. *Photogr. Eng.*, 28, pp. 625-632, 1962.
- E. Mikhail (1963): Use of Two-Directional Triplets in a Sub-Block Approach for Analytical Aerotriangulation. *Photogr. Eng.*, 29, pp. 1014-1024, 1963.
- D. Mumford (2000): The Dawning of the Age of Stochasticity, in Arnold, V. (ed.) et al., *Mathematics: Frontiers and Perspectives*. Providence, RI: American Mathematical Society (AMS). pp. 197-218, 2000.
- D. Mumford and J. Shah (1989): Optimal Approximation by Piecewise Smooth Functions and Associated Variational Problems. *Comm. Pure Appl. Math.*, 42: pp. 577-685, 1989.
- D. Nistér (2003): An Efficient Solution to the Five-Point Relative Pose Problem, In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Volume 2, pp. 195-202, 2003.
- S. Osher, J. A. Sethian (1988): Fronts Propagating with Curvature Dependent Speed: Algorithms Based on Hamilton-Jacobi-Formulation, *J. Comput. Phys.* 79, pp. 12-49, 1988.
- B. Triggs, P. McLauchlan, R. Hartley, A. Fitzgibbon (2000): Bundle Adjustment – A Modern Synthesis, in Triggs, Zisserman, Szeliski: *Vision Algorithms: Theory and Practice*, LNCS 1883, pp. 298-372, Springer, 2000, <http://lear.inrialpes.fr/pubs/2000/TMHF00/>, last visited 20. 7. 2009