Christoph Musik

Ground Truth Studies: A Socio-Technical Framework

103 - Recent Advances in Multimedia Processing, Organization and Visualization beyond Domains and Disciplines

Abstract

By proposing the analytical reflection framework of 'Ground Truth Studies', this paper stresses the importance and significance of computer vision Ground Truth and its construction process as a constituting socio-technical element. The framework of Ground Truth Studies is a specific form of Socio-Technical Integration Research (STIR) aimed at bringing together different actors from the fields of computer sciences, social sciences, the humanities, and the arts. As such, it is a concrete laboratory-based manifestation of what is referred to as Responsible Research and Innovation (RRI).

Keywords:

Ground Truth, Image Processing Algorithms, Training Images, Responsible Innovation, Socio Technical Integration Research (STIR)

1. Introduction

In the age of Big Data, coping with information relates in particular to the ever-increasing amount of visual data that is constantly produced in and about the physical world. Optical information deriving from movements of bodies and non-human entities (e.g. objects, cars, planes) needs to be understood in terms of image processing and computer vision algorithms in order to monitor, control, take care of, track, and manage people and objects. Since image processing algorithms (IPAs) are increasingly becoming powerful societal actors and decision-makers in the course of the greater socio-technical transformation processes of digitalisation and automation, it is important to understand exactly and reflect carefully on the production, processing, and interpretation of digital images by algorithms where the semantic interpretation element plays the central role.

In order to achieve this objective, this paper proposes the socio-technical framework of Ground Truth Studies as both an analytical and a reflective framework that refers specifically to the field of computer vision and visual pattern recognition. Furthermore, the framework of Ground Truth Studies is conceptualised as a means of bringing together and encouraging a range of actors from a range of different areas and fields to work and collaborate as a specific form of Socio-Technical Integration

Research (STIR) (Schuurbiers 2001). As such, this serves as a concrete manifestation of what is referred to as Responsible Research and Innovation (RRI).

The paper is organised as follows: first, the theoretical background of the framework of Ground Truth Studies is presented by referring to Studies of Classification and Standardisation that have been influenced in particular by sociology and the interdisciplinary fields of Surveillance Studies as well as by Science and Technology Studies (STS). The second part of the paper illustrates the importance and significance of Ground Truth data as a constituting socio-technical element by referring to my own empirical research on IPAs in the areas of face recognition, facial expression recognition, behaviour detection, and fall detection. Finally, the framework of 'Ground Truth Studies' is presented as a specific application of so-called 'Socio-Technical Integration Research' (STIR) (Schuurbiers 2001).

2. Classification and Standardisation

Image Processing Algorithms (IPAs) are fundamentally based on 'situated' classification and standardisation practices. Therefore, IPAs and their construction pose "sharp questions for democracy", because they "may (then) come to function as an alternative to expert authority" (Timmermans / Epstein 2010: 70-71), which might be contained as such "in rules and systems rather than in credentialed professionals" (ibid. 71).

The central area where these 'situated' classification and standardisation practices emerge in computer vision lies in the sociotechnical construction of 'Ground Truth'. The IPA Ground Truth and its creation is a crucial and mainstay societal element, as it defines and standardises what is perceived to be real and true. It can be regarded as the production of an 'interpretation template', or, under specific circumstances, as the production of a 'truth template' that is the basis for all image interpretation done by IPAs. It is clear that those involved in these construction processes exert power, whether intentionally or not, on account of their being able to decide what counts as relevant knowledge in each and every particular case (Forsythe 1993). Thus, they are not only in a position to decide and define what is real and what is true in the world, but are also simultaneously in a position to decide what is to be defined as desirable and undesirable, what is good and what is bad. It is then a way of "constructing uniformities across time and space through the generation of agreed-upon rules" (Timmermans / Epstein 2010: 71). The problem is that these "agreed-upon rules" are very particular and situation-dependent and might ultimately contain a wide array of tacit values and assumptions that represent the viewpoints of particular individuals. Coding never occurs in an objective or neutral way, but is embedded in specific, socially situated practices and actions. Bowker and Star (2000) see computer software in many ways as "frozen organizational and policy discourse", in which policy is coded into software. In this view, software, like technology, is "society made durable" (Latour 1991). What is problematic with this view is that "the exercise of this power is to some extent invisible" (Forsythe 1993: 469). This means that the engineers' specific (e.g. male, western etc.) 'situated' view, with all its tacit values and assumptions, is being "black-boxed" (Latour 1999: 304) and thus stabilised over time. Nevertheless, it is perceived by the user of such a system as being ‘correct’ and ‘true’ in every sense.

3. The Societal Significance of the Ground Truth

By proposing the analytical reflection framework of ‘Ground Truth Studies’ in order to address the quiet ‘politics of classification’ (Bowker / Star 2000: 195ff.), the importance and significance of Ground Truth data as a constituting socio-technical element will be explained in this part of the paper. This will be done by referring to my own social scientific empirical research on IPAs and Ground Truth construction processes in the areas of face recognition, facial expression recognition, behaviour detection, and fall detection. The empirical research presented here refers to many years of analysis of the field of computer vision. It is based in particular, on ethnographic fieldwork in and around an Austrian computer vision laboratory in 2011; on the output of an inter- and trans-disciplinary research project within the Austrian security research scheme KIRAS in which I was involved as a contract researcher between 2009 and 2010; and on many formal and informal interviews and conversations with computer scientists from around the world.

3.1 What is Ground Truth?

In order to explain the significance of Ground Truth as a constituting socio-technical element, it is important to first clarify what is actually meant by Ground Truth in this context. The basis for teaching a machine or a computer to see and to recognise is the engineering of this so-called ‘Ground Truth’ (sometimes also referred to as ‘Ground Reality’) of what a specific entity of interest might eventually look like. How does it work? A computer vision scientist working on facial expression recognition explained it to me in the following way:

> “ …you give the machine example data to train from. So for instance if you want a machine to recognise a specific person then you show the machine images of this person and you tell the machine that this image shows that person. You give the correct answer already in the training phase. If you want to recognise laughing or fear or whatever, you show the machine images of laughing or afraid persons and you tell the machine these images show laughing or afraid persons. And so the machine can recognise it later. But in the training phase this information has to be given and this is called Ground Truth.”

Constructing Ground Truth, which is the essential basis for any image processing and pattern recognition, is a highly complex and above all time-consuming process, which stands quite unpopular in the everyday practice of computer vision. As I was told several times, this is also the reason why this process is often outsourced to students or interns. Nevertheless, when considering the importance of Ground Truth for future computer vision detections and their viability, special attention to and careful reflection on how knowledge is conceptualised and processed and what this implies is needed, for example, in terms of different types of biases and what could be called “smart” discrimination in the wake of these biases. As an illustration of this, in their introductory book on image processing and analysis, Sonka et al. (2008) explain the challenges and possibilities of computer vision using the example of a cow. They describe that following a ‘training phase’ in which the system is taught what a cow might look like in various poses, a model of a cow in motion can be derived. In consequence:

> “these models could then be fitted to new (‘unseen’) video sequences. Crudely, at this stage anomalous behaviour such as lameness could be detected by the model failing to fit properly, or well.” (ibid.2).

One central assumption in this statement is that of similarity (van der Ploeg 2011: 30) and “lumping” (Zerubavel 1996). In the context of “lumping”, Zerubavel brings in the term “island of meaning” - a cluster of things (here: cows) that are regarded as “more similar to one another than to anything outside the cluster” (ibid. 422). Sonka et al. (2008) point out that when the system is taught what a cow might look like, it is assumed that there is only one universal cow. One look at the global databank of animal genetic resources shows that there are 897 reported regional cattle breeds, 93 regional transboundary cattle breeds, and 112 international transboundary cattle breeds (FAO 2007: 34ff.). This means that there is certainly not one universal kind of cow, but in fact a reported total of 1102 different cattle breeds worldwide. This example makes clear what Forsythe’s insight into the brittleness and narrowness of background knowledge that is taken for granted (Forsythe 1993: 467) means for computer vision. In order to teach the computer what something, e.g. a cow, looks like, the human computer scientist has to give example data about the object of interest in a training phase. For instance, if the computer scientist is based in Austria and predominantly uses images of cows showing the most widespread Austrian ‘Fleckvieh’ cattle in order to teach the computer how cattle generally look, the possibility of recognising the 1101 other breeds such as, for example, the Ugandan Ankole cattle might be lower and thus, the algorithm excludes all but Austrian Fleckvieh. What has occurred is an example of the ignorance of intracluster differences (Zerubavel 1996: 423). In such a case, Austrian Fleckvieh cattle would be the standard and norm of what a cow looks like, performing a specific stereotype appearance of a real cow. The same can be applied to things like object detection or the detection of suspicious behaviour, for example, to recognise criminal or terrorist attacks. Yet what does suspicious behaviour actually look like in concrete situations, and what distinguishes it from non-suspicious, normal behaviour? What is the knowledge base for all this?

These examples help to understand that constructing the ground truth in computer vision laboratory work is based on culturally situated classification and standardisation practices that do not come into being arbitrarily, and that do not always rely on an objective, neutral, technical or natural foundation. Ground Truth and its concomitant creation then, represent a crucial and fundamental societal element, as it defines and consistently standardises what is perceived as real and true, and what is not.

3.2 Ground Truth Studies: Basic Questions

What follows from these empirical insights are important basic questions for Ground Truth Studies: what kind of knowledge or visual expertise is used in order to produce the respective Ground Truths? Is it more formalised and explicit or less formalised, tacit knowledge? Is it based on expert views or on everyday common sense? In this process, it is crucial to consider what aspects influence, characterise and are embedded in the respective Ground Truths. Why exactly were these aspects chosen and of significance? What proof is given that the applied characteristics are real evidence for the specific domain of scrutiny? For example, referring to a system of automated fall detection, a question of interest lies in the situation of how it can be proved that the relationship between a straight line representing the human body and a plane representing a detected floor area indicates, for example, whether a person has had a significant fall. All in all, it should be clear which specific and particular version of reality and truth has been transferred to and manifested in each respective Ground Truth. Once it has been formalised and determined, the question can be asked if there is still room to either use the respective Ground Truth for image comparison, or to allow alternative (human) views, e.g. at another place or at another point in time.

As far as the construction of a Ground Truth is concerned, the selection and application of training images also figure prominently, as they influence and finally define the particular Ground Truth model. Important questions here are: Why are specific images chosen? How are these images constituted? What sources do they come from and in what way are they used to give evidence of an entity?

Finally, in the course of Ground Truth evaluations, bias studies are an important means for the analysis of possible discrimination and new types of digital divide. Introna and Wood demand "bias studies" in the context of their analysis of the politics of face recognition technologies, especially when implemented in CCTV systems (Introna / Wood 2004). One of their central results was that facial recognition algorithms seem to have a systemic bias: men, Asian and Afro-American populations, as well as older people are more likely to be recognised than women, white populations and younger people (ibid. 190). A consequence of this bias could be, for example, that those with a higher possibility of being recognised are those with a higher probability of scrutiny or of setting off an alarm. As a consequence of these findings, the question is raised of what can be done to limit biases (ibid. 195). As many biases seem to be inscribed into IPAs unintentionally, it is important at least to analyse biases once they are in operation, but it is advisable to analyse the relation of Ground Truth construction and systematic bias at an early stage, before IPAs are implemented in operating systems. Because most IPA systems in operation are inaccessible for external scrutiny, another possibility for gaining information about biases is an obligation to investigate biases and publicise the results of bias studies before a system affects people in a negative way.

4. Ground Truth Studies as Socio-Technical Integration Research (STIR)

The framework of 'Ground Truth Studies' is presented here as a specific applied form of 'Socio-Technical Integration Research' (STIR) (Schuurbiers 2001). STIR can be defined as "any process by which technical experts account for the societal dimensions of their work as an integral part of this work" (Fisher / Maricle 2015: 74). As such, it is an important element of Responsible Research and Innovation (RRI) (Stilgoe et al. 2013), focusing on laboratory-centred integration and intervention of the Social Sciences and Humanities (SSH). The socio-technical framework of Ground Truth Studies explicitly suggests early collaboration of computer scientists, designers, engineers, SSH, and other societal actors such as artists working on or with IPAs. The involvement of other societal actors in IPA research and development might help to give computer vision more solid grounding, as, when it comes to societal implementation, aspects of friction are outspokenly questioned from the very beginning. Here it is important to comment on the relation between SSH scientists and laboratory practitioners. It is not the case that computer scientists have a general ‘reflective deficit‘ and social scientists are more reflective. Rather, it is the case that the knowledge held by social scientists or other actors could complement computer scientists´ knowledge through interdisciplinary collaboration. In this regard, diversity is key to creating new understandings of socio-technical innovation processes (Felt 2014) because human vision is situated and particular (Burri 2013). It is thus important to consider and make use of a great variety of situated and particular views that potentially contradict the situated and specific view of computer scientists. Involving other people with other views could therefore help to inscribe more diversity (and in this way, more democracy) into IPAs and in turn, it could help to reduce – yet fall short of fully eliminating – influential semantic gaps. Finally, it might also be important to warn against enrolling SSH in laboratory practices “as strictly symbolic and superficial partners without influence over research and innovation activities” (Gjefsen / Fisher 2014: 428). As a consequence, integrating SSH into laboratory practices must occur on an equal footing (Felt 2014). In this sense, Ground Truth Studies as Socio-Technical Integration Research are designed to accept and integrate different forms of knowledge that, when in a smart combination, lead to solutions enhancing societies in desirable ways.

Literaturliste/Quellenverzeichnis:

Bowker, Geoffrey C./Star, Susan Leigh (2000): Sorting things out. Classification and its consequences. Cambridge & London: The MIT Press.

Burri, Regula Valérie (2013): Visual Power in Action: Digital Images and the Shaping of Medical Practices. In: Science as Culture 22 (3): 367-387.

Felt, Ulrike (2014): Within, Across and Beyond – Reconsidering the Role of Social Sciences and Humanities in Europe. In: Science as Culture 23 (3), 384-396.

Fisher, Erik /Maricle, Genevieve (2015): Higher-level responsiveness? Socio-technical integration within US and UK nanotechnology research priority setting. In: Science and Public Policy 42 (1), 72–85.

Forsythe, Diane E. (1993): Engineering Knowledge: The Construction of Knowledge in Artificial Intelligence. In: Social Studies of Science 23 (3), 445-477.

Gjefsen, Mads Dahl/Fisher, Erik (2014): From Ethnography to Engagement: The Lab as a Site of Intervention. In: Science as Culture 23 (3), 419-431.

Introna, Lucas D./Wood, David (2004): Picturing Algorithmic Surveillance: The Politics of Facial Recognition Systems. In: Surveillance & Society 2 (2/3) 177–198.

Latour, Bruno (1991): Technology is society made durable. In: Law, John (ed.): A sociology of monsters: essays on power, technology and domination. London & New York: Routledge, 103-131.

Latour, Bruno (1999): Pandora’s hope: essays on the reality of science studies. Harvard University Press.

Schuurbiers, Dan (2011): What happens in the Lab: Applying Midstream Modulation to Enhance Critical Reflection in the Laboratory. In: Science and Engineering Ethics 17, 769-788.

Sonka, Milan/Hlavac, Voclav/Boyle, Roger (2008): Image Processing, Analysis, and Machine Vision. CL-Engineering. 3rd edition.

Stilgoe, Jack/Owen, Richard/Macnaghten, Phil (2013): Developing a framework for responsible innovation. In: Research Policy 42, 1568–1580.

Timmermans, Stefan/Epstein, Steven (2010): A World of Standards but not a Standard World: Toward a Sociology of Standards and Standardization. In: Annual Review of Sociology 36, 69–89.

Van der Ploeg, Irma (2011): Normative Assumptions in Biometrics: On Bodily Differences and Automated Classifications. In: Simone van der Hoof & Marga M. Groothuis: Innovating Government. Normative, Policy and Technological Dimensions of Modern Government. The Hague: TMC Asser Press, 29–40.

Zerubavel, Eviatar (1996): Lumping and Splitting: Notes on Social Classification. In: Sociological Forum 11 (3): 421-433.