The genomics revolution is changing medicine; it could transform visual surveillance and face recognition next

Despite decades of research, automatic visual surveillance systems have not delivered the results initially promised to the public. Monitoring of the videos produced by camera networks remains largely the task of security officers. A main issue is the way those visual surveillance systems are conceived. Usually, they are developed first in laboratories where high resolution cameras positioned at human height sit on robust tripods, light conditions are fully controlled, the number of individuals visible in the field of view is small and the potential for occlusions is limited. In such environment, standard video analytics approaches perform extremely well. However, when such systems are deployed outdoor, ‘refinements’ become necessary. Special filters need to be developed to take care of an ever growing list of ‘inconveniences’, such as shadows, brightness and weather condition changes, scene and view angle variability, camera vibrations and the large range of distances between subjects and cameras. Although the systems’ performance suffers, they still prove to be useful in specific scenarios, especially when supported by human operators. However, once cameras stop being stationary, the complexity of the task increases significantly. While PTZ (Pan, Tilt, Zoom) cameras have very constrained motions which can be ‘compensated’ by including camera motion models, the case of free moving cameras appear essentially intractable. Therefore, with the increased usage of such cameras by police forces, e.g. in-car, body-worn and drone-based cameras, the model of incremental updating of a system initially designed for a control environment has reached its limits. Not only does it lead to the building of ever more complex and inconsistent systems, but it results in increasing performance degradation to the point of becoming totally inadequate. Consequently, a new strategy is required where, instead of attempting to control the huge number of parameters which may affect a scene as captured by a camera, scene variability is considered as the expected norm rather than a nuisance to eliminate.

Since evolution is the main process behind life's diversity, variability is at the core of the analysis of the organisms’ genetic material, i.e. the field of genomics. With international efforts such as the Human Genome Project [1], which sequenced the 3 billion DNA characters of the human genome, thousands of complete genomes are now available and this number is increasing at an exponential pace. Due to their enormous potential for human health and medicine, and the inability of standard algorithms to process such data, the pharmaceutical industry, biotech companies and publicly-funded institutions have spent vast amounts of money and resources to develop the specialised software and computing facilities able to handle the specificity, the inherent variability and the sheer size of genomics data. With international research organisations, such as the European Bioinformatics Institute [2], which deliver mature and powerful software services to millions of scientists, the genomics revolution is becoming a reality already delivering personalised medicine for conditions such as cancer [3]. While variations of the BRCA1 gene, so called

Angelina Jolie gene, may indicate an 80% risk of developing breast cancer, the identification of the specific mutation combined with classical risk factors and mammographic density can reveal an individual reduced risk of 30% [4].

The vide-omics paradigm

For the last 10 years, my research group at Kingston University, London, has investigated how those genomics approaches could benefit visual surveillance by putting variability at the centre of its data analysis. This has led to the recent publication of a genomics-inspired paradigm for video analysis - vide-omics [5]. It is important to note that this concept is fundamentally different from 'genetic algorithms' which were designed for optimisation and searching complex spaces. In the vide-omics paradigm, a video is considered to be simply a set of temporal measurements of a scene which is in constant evolution. As a result, there is no constraint regarding camera motion, scene structure or types of changes. Using a genetics analogy, the background of a scene described by a set of videos can be seen as the common ancestor from which all videos' frames have diverged from. As a consequence, video analysis becomes equivalent to the interpretation of mutations which have been detected on each frame.

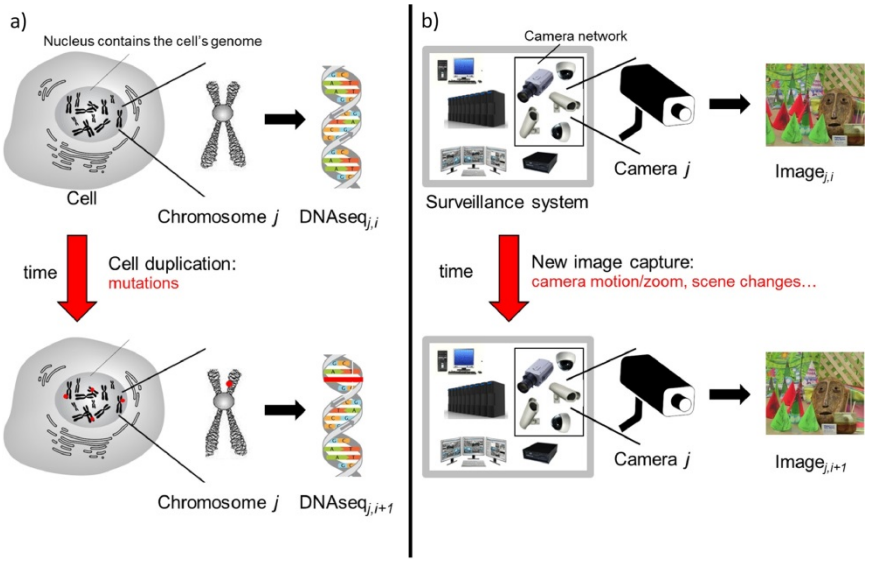

Figure 1: Analogy between (a) cell duplication and (b) video capture by a surveillance system.

In genetics, DNA is affected by five main types of mutations - substitution, insertion, deletion, duplication and transposition – where DNA characters are respectively replaced, added, removed, duplicated or moved to another location. Remarkably those mutations have equivalence in the visual surveillance world where pixel values are modified according to similar processes:

- Substitution: pixel value variation due to sensor noise, change in camera gain, change in scene illumination.

- Insertion: change in camera angle and/or position revealing previously occluded data, appearance of more details in a camera's field of view (zoom in), change of field of view (camera motion or zoom out) leading to the appearance of new objects.
- Deletion: change in camera angle and/or position introducing new occlusions, loss of details in field of view (zoom out), disappearance of objects from field of view after camera motion or zoom in.
- Duplication: scene area seen by overlapping cameras.
- Transposition: foreground object motion.

Therefore, genomics algorithms originally developed to analyse mutations in strings of DNA characters can relatively easily be adapted to handle strings of pixels. Recent implementations of the vide-omics paradigm in challenging scenarios have established its validity. First, it was demonstrated that dense pixel matching, an essential step in stereo correspondence, could be performed on unrectified or non-linearly distorted images using neither camera nor lens distortion models [6]. Second, it was shown that foreground extraction could be achieved on videos captured by a freely moving camera, performing consistently despite a variety of camera motions and scene structures [5]. While there is still a lot of scope for progress, a new avenue has now been opened, providing new possibilities for addressing the most demanding computer vision tasks.

Future application to facial recognition

Despite noticeable success of facial recognition systems such as those operating at e-Passport gates or offering automatic name tagging on photos shared on social networks [7], one is also regularly reminded of their limitations. For instance, the embarrassing failure of the system recently tested by the Metropolitan police at the Notting Hill carnival in 2017 was largely reported in the media [8]. In order to understanding those somehow conflicting reports about facial recognition performance, one needs to classify biometric scenarios according to the subject's attitude. Indeed, an individual can be either cooperative, i.e. making efforts so that the system recognise them, non-cooperative, typically they are not aware of the presence of a facial recognition system, or uncooperative, i.e. trying to prevent a system to recognise them by, for example, covering their face or looking down. While the e-gate and the social network applications chiefly deal with cooperative individuals - in a social situation people tend to look towards the camera when a photo is taken -, visual surveillance always operate with non-cooperative or uncooperative subjects. In addition, unlike the other systems, visual surveillance software generally have to handle unconstrained environments. As a consequence, automatic facial recognition applied to visual surveillance remains very much a research topic.

Conversely, some extraordinary individuals, the so-called super recognisers, have the ability of accurately identifying individuals captured by CCTV cameras. Could one learn from the mental process they employ to design more effective facial

recognition software? Studies have revealed that, not only do super recognisers have excellent photographic memory, but they also have the capability to extrapolate from several viewpoints [9]. This suggests that algorithms relying on matching features between image pairs are unlikely to be able to replicate their approach.

Figure 2: Protein multiple alignment of HIV-1 envelope sequences [10]. Highly conserved characters are highlighted by colours showing their chemical property. Dashes represent 'missing' characters.

Fortunately, systems based on the vide-omics paradigm have that potential. All variations of a human face can be interpreted as the product of mutations applied to a canonical face, their common ancestor. In genomics, the model of a common ancestor is commonly inferred by performing first a multiple sequence alignment of biological sequences belonging to related species [11]. That alignment aims at finding the optimal correspondences between all the sequences' individual characters across all sequences while being consistent with possible mutation events. Then, a model of the multiple sequence alignment can be generated using techniques, such as profile Hidden Markov Models, so that eventually members of the family defined by those species can be identified [12]. A similar methodology could be applied to build a statistical model of the canonical face which would have produced all known face images of a given individual. Indeed, instead of 'aligning' all those faces using global geometric transformations, the multiple alignment needs to be performed at the pixel level. By doing so, differences due to view orientations are taken into account without explicitly attempting to build a 3D model of the face: for example, areas hidden by the nose on certain views would be treated as pixel insertions when visible on other views. Similarly, illumination variations and addition of glasses or a beard would be modelled as pixel substitutions. Finally, the various facial expressions would conduct to individual pixel shifts, the modelling of which would represent quantitatively the locally relatively static areas - associated to bone and cartilage structures - and the more dynamic ones moved by the facial muscles. Eventually, a face of interest could be compared with a database of face models. The issue would then be to decide if the best hits are significant. As the 'infinite monkey theorem' illustrates the probability of finding 'good hits' in a database increases with its size, and already the UK polices hold millions of faces [13]. Again, the field of genomics would be able to help since, for already many years, it has had to deal with databases commonly containing hundreds of millions of entries. Thus, in order to evaluate the relevance of a given alignment, tools provide the number of random alignments which are expected to have the same quality (E-value), given the

sizes of the database and the entry sequence [14]. As a consequence, any hit with an E-value above 1 can be interpreted as obtained by pure chance and the lower the E-value the more confident one can be of the significance of an alignment. The adaptation of that mathematical framework to face recognition may go a long way to reducing the number of ‘erroneous arrests’ which give a bad name to visual surveillance [8].

Nowadays, the public expects to be protected by automatic visual surveillance systems able to accurately identify faces from CCTV footage. This expectation is further exacerbated by the numerous television crime and espionage thrillers which present that technology as a standard, generally infallible, tool available at the press of a button. Bridging the gap between reality and fiction will take time. Hopefully, the adoption of the vide-omics paradigm will make it shorter.

References

[1] I.H.G.S. Consortium, et al., Initial sequencing and analysis of the human genome, Nature, 409: 860-920, 2001

[2] C.E. Cook, M. T. Bergman, R. D. Finn, G. Cochrane, E. Birney and R. Apweiler, The European Bioinformatics Institute in 2016: Data growth and integration, Nucleic Acids Research, 44(D1): D20-6, 2015

[3] B.C. Yoo, K.-H. Kim, S. M. Woo and J. K. Myun, Clinical multi-omics strategies for the effective cancer management, Journal of Proteomics, doi.org/10.1016/j.jprot.2017.08.010, 2017

[4] J. Walrond, Gene test 'narrows down breast cancer risk', BBCNews, 8 October 2017, Accessed November 2017. http://www.bbc.co.uk/news/health-41503013

[5] I. Kazantzidis, F. Florez-Revuelta, M. Dequidt, N. Hill and J.-C. Nebel, Vide-omics: A genomics-inspired paradigm for video analysis, Computer Vision and Image Understanding, doi.org/10.1016/j.cviu.2017.10.003, 2017

[6] J. Thevenon, J. Martinez del Rincon, R. Dieny and J.-C. Nebel, Dense Pixel Matching Between Unrectified And Distorted Images Using Dynamic Programming, International Conference on Computer Vision Theory and Applications, 24-26 February, Rome, Italy, 2012

[7] M. Crampes, J. Oliveira-Kumar, S. Ranwez and J. Villerd, Visualizing social photos on a Hasse Diagram for eliciting relations and indexing new photos, IEEE Transactions on Visualization and Computer Graphics, 15 (6): 985-992, 2009

[8] A. J. Martin, Police facial recognition trial led to 'erroneous arrest', Sky News, 7 September 2017, Accessed November 2017. https://news.sky.com/story/police-facial-recognition-trial-led-to-erroneous-arrest-11013418

[9] T. Ring, Humans vs machines: the future of facial recognition, Biometric Technology Today, pp. 5-8, April 2016

[10] A. Abecasis, A. Vandamme and P. Lemey, Sequence Alignment in HIV Computational Analysis, pp. 2-16 in HIV Sequence Compendium 2006/2007, Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM. LA-UR 07-4826, 2007

[11] C. Notredame, D. Higgins and J. Heringa, T-coffee: a novel method for fast and accurate multiple sequence alignment, Journal of Molecular Biology, 302(1): 205-217, 2000

[12] R. D. Finn, J. Clements and S. R. Eddy, HMMER Web Server: Interactive Sequence Similarity Searching, Nucleic Acids Research, 39: W29-37, 2011

[13] A. J. Martin, Police hold more than 20 million facial recognition images, Sky News, 24 August 2017, Accessed November 2017. https://news.sky.com/story/police-hold-more-than-20-million-facial-recognition-images-11001479

[14] W.R. Pearson, Comparison of methods for searching protein sequence databases, Protein Science, 4:1145-1160, 1995.

About the author

Dr Jean-Christophe Nebel is an Associate Professor in the School of Computer Science and Mathematics at Kingston University, London. He leads research in pattern recognition applied to computer vision and bioinformatics. His contributions include the development of a 3D television studio, a stochastic context free grammar-based framework able to handle the protein alphabet, a dimensionality reduction method that can model any concept represented by multivariate sequences, and a genomics-inspired paradigm for video analysis (vide-omics).