Computer Vision and Image Understanding 166 (2018) 28–40

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Vide-omics: A genomics-inspired paradigm for video analysis

Ioannis Kazantzidis*,[a], Francisco Florez-Revuelta[b], Mickael Dequidt[c], Natasha Hill[d], Jean-Christophe Nebel[a]

[a] School of Computer Science & Mathematics, Faculty of Science, Engineering and Computing Kingston University, London, Kingston-Upon-Thames, KT1 2EE, UK
[b] Department of Computer Technology, University of Alicante, P.O. Box 99, E-03080 Alicante, Spain
[c] Service Ressources Informatiques et Communications, Ifremer, Brest, Centre Bretagne, 29280 Plouzane, France
[d] School of Life Sciences, Pharmacy & Chemistry, Faculty of Science, Engineering and Computing Kingston University, London, Kingston-Upon-Thames, KT1 2EE, UK

ARTICLE INFO

Keywords:
Computer vision
Freely moving camera
Genomics
Foreground detection
Segmentation
Scanlines

ABSTRACT

With the development of applications associated to ego-vision systems, smart-phones, and autonomous cars, automated analysis of videos generated by freely moving cameras has become a major challenge for the computer vision community. Current techniques are still not suitable to deal with real-life situations due to, in particular, wide scene variability and the large range of camera motions. Whereas most approaches attempt to control those parameters, this paper introduces a novel video analysis paradigm, 'vide-omics', inspired by the principles of genomics where variability is the expected norm. Validation of this new concept is performed by designing an implementation addressing foreground extraction from videos captured by freely moving cameras. Evaluation on a set of standard videos demonstrates both robust performance that is largely independent from camera motion and scene, and state-of-the-art results in the most challenging video. Those experiments underline not only the validity of the 'vide-omics' paradigm, but also its potential.

1. Introduction

Introduction of cameras in public places has been associated with the promise that they would contribute to a safer and more secure society. However, the amount of generated video is such that that pledge can only be delivered if CCTV operators are supported by video analysis tools which could identify, detect or, at least, suggest objects or actions of interest. Although state-of-the-art video processing algorithms have been the product of extensive work for decades, current approaches are still not sufficient to deal with the very wide range of data exhibited by CCTV imagery in real-life situations. Whereas most methods attempt to control the huge number of parameters affecting a scene, an alternative strategy would be to design methodologies addressing variability at their core. This motivates the proposal of a novel video analysis paradigm, 'vide-omics', founded on the principles of genomics where variability is the expected norm rather than an inconvenience to control.

Analogies can be drawn between genomics data and images in terms of structure and evolution. Similarly to an image which can be encoded as a set of pixel strings, genetic material has essentially a linear digital structure which is represented by strings of millions of characters, called sequences. Those sequences evolve over time through mutations of single and group of characters. Likewise, a continuous video can be interpreted as the capture of a single image evolving through time. Thus, video analysis could be addressed by detecting and quantifying image mutations over time. A benefit of the proposed paradigm is that it does not impose any constraint on the way videos are captured. As a consequence, it should be able to handle videos recorded by freely moving cameras. The 'vide-omics' paradigm aims at not only providing a novel way of describing video data where variability is the norm, but also to harvest the mature methodologies used for genomics analysis in order to apply them to video processing.

The objectives of this paper are, first, to introduce the video analysis paradigm, 'vide-omics', and, second, to provide a proof of concept by applying it to foreground extraction from videos captured by freely moving cameras. After introducing relevant genomics concepts and exploring their previous exploitation in computer vision, a review of the state of the art for foreground segmentation in the context of freely moving cameras is provided. Then, the 'vide-omics' paradigm is presented and its application to foreground extraction is described. Finally, it is evaluated on a set of standard videos recorded by freely moving cameras and performance is discussed.

* Corresponding author.
E-mail address: ikazant@kingston.ac.uk (I. Kazantzidis).

http://dx.doi.org/10.1016/j.cviu.2017.10.003
Received 3 November 2016; Received in revised form 29 September 2017; Accepted 8 October 2017
Available online 10 October 2017

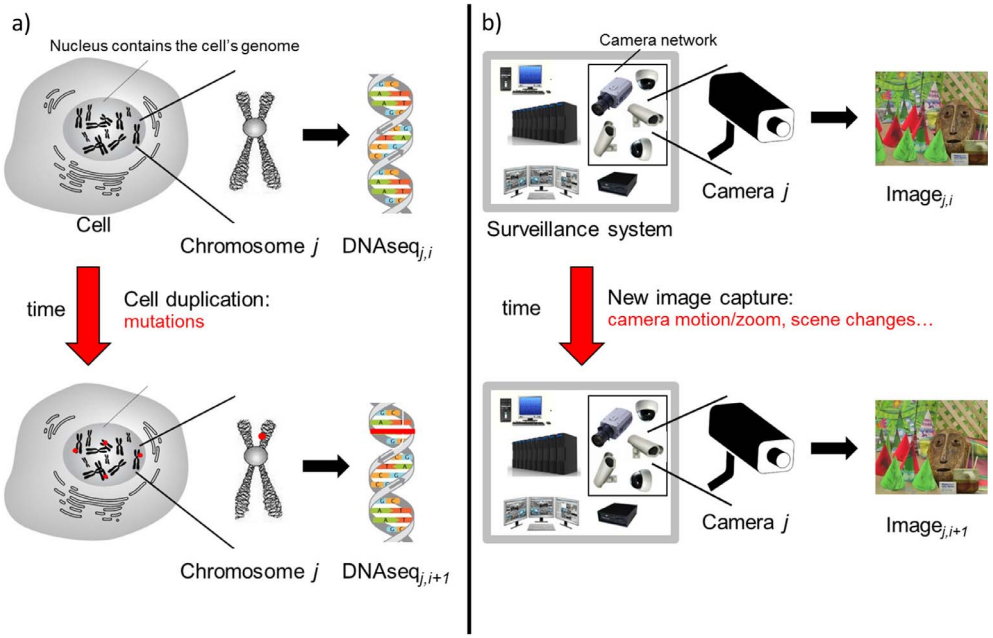

Fig. 1. Analogy between (a) Cell duplication and (b) video capture by a surveillance system.

2. Related work

2.1. Relevant genomics concepts

Genomics is the field of genetics which combines experimental techniques and computational approaches called bioinformatics, to sequence, assemble and analyse the genetic material of organisms, i.e. their genome. In the living cell, the genome is stored in long double chains of deoxyribonucleic acid (DNA) which are packed in individual chromosomes, see Fig. 1a. Those chains total from 0.1M in some bacteria to Gigas of building blocks nucleotide pairs - in high-order organisms. DNA is made of four types of nucleotides: adenine (A), cytosine (C), guanine (G), and thymine (T). During the process of cell duplication, two identical copies of DNA are produced. Although that process is highly accurate, mistakes still occur with an error rate of the order of 10^{-9} McCulloch and Kunkel (2008); they are the basis for organism evolution and genetic disorders. Types of replication errors are varied and include: insertions, substitutions, deletions, duplications and transpositions, where individual or groups of nucleotides are respectively added, replaced, deleted, duplicated and moved within or between DNA chains.

With international efforts such as the Human Genome Project Consortium et al. (2001), 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. Their analysis has required not only the applications of conventional data mining and pattern recognition approaches, but also the development of completely novel techniques to handle the specificity and sheer size of genomics data Sebastiani et al. (2003); Zhang (2007). With the expectation that deciphering the human genome will result in dramatic improvement of health, the international community has required from bioinformatics to produce fast, efficient and robust computational techniques tailored to genome analysis Fernandez-Suarez and Birney (2008); Medvedev et al. (2009). As a consequence, nowadays bioinformatics organisations, such as the European Bioinformatics Institute, deliver mature and powerful tools which serve millions of scientists Brooksbank et al. (2014).

Since genomics relies on DNA sequence comparisons to infer evolutionary relationship, predict the sequence of a common ancestor and provide function annotations, numerous bioinformatics tools have been developed to find optimal character correspondences or alignment - between a set of sequences (multiple sequence alignment). Most of them, including the currently most popular ones Altschul et al. (1997); Mackey et al. (2002); Notredame et al. (2000); Thompson et al. (1994), rely on some derivation of the Needleman–Wunsch algorithm Needleman and Wunsch (1970), which was the first effective and automatic method to produce an exact solution to the global alignment of two sequences.

2.2. Exploitation of bioinformatics in computer vision

In the last few years, a few research groups have had a common objective: the exploitation of bioinformatics ideas and approaches for pattern recognition problems in computer vision. Initially, analogy between DNA sequences and image sequences was explored to take advantage of DNA sequence comparison approaches to compare videos. Riedel et al. (2008) adapted the Smith–Waterman local alignment approach from bioinformatics (Smith and Waterman, 1981) to measure video similarities for activity recognition. The Video Genome Project at Technion went further in their analogy by proposing to treat the task of identifying and synchronising different versions of a video as the alignment of two mutated sequences sharing a common ancestor. Their approach relies on local alignments of video sequences, where each frame is represented by a histogram of quantized salient point descriptors. Despite encouraging performance (Bronstein et al., 2010), there is no evidence that further work was carried on based on that concept. Bicego et al. (2015); Bicego and Lovato (2012); 2016); Lovato et al. (2014) from the University of Verona have proposed encoding 2D and then 3D shapes as a biological sequence so that actual bioinformatics comparison tools could be used for shape recognition and classification. Their very competitive results have validated their approach. Finally, Nebel et al. have made a sustained effort in addressing various tasks of stereo matching as sequence alignment problems: finding correspondences between scanlines (Dieny et al., 2011), a scanline and a curve in an unrectified and distorted image (dos Santos-Paulino et al., 2014; Thevenon et al., 2012) and eventually implementations on various low-cost and low-complexity embedded devices (Madeo et al., 2014). All those applications support the idea that bioinformatics research has a lot to offer to the pattern recognition communities and to computer vision in particular. Here it is proposed to go beyond opportunistic exploitation by offering a new paradigm for video processing: 'vide-omics'. In such a framework, a video is seen as the record of

a scene evolving through time so that its analysis can be performed by detecting and quantifying scene mutations over time.

2.3. Background/foreground segmentation for freely moving cameras

In computer vision, background/foreground (B/F) segmentation refers to the process of discriminating between moving or foreground objects and static objects within a video. Common challenges include noisy images, camera jitter, illumination changes, shadows, physical motion in background, e.g. moving tree leaves, and zooming (Toyama et al., 1999). Since further complexity is added when the camera is not fixed, the computer vision community has focused mainly on stationary camera set ups for which more than 300 methods have been proposed (Bouwmans, 2011). Addressing that segmentation task for freely moving cameras has become more and more important with the development of applications associated to ego-vision systems, smart-phones, and autonomous cars. Currently, methodologies are divided into two main categories: camera-based models that attempt to compensate for camera's motion and approaches that analyse pixel motions. While camera-based models relies on homography, epipolar geometry or a combination of both, pixel motion analysis either consider long-term trajectories or per frame dense pixel motion. When camera motions are limited to pan, tilt and zoom (PTZ), the standard approach is to create an image mosaic (Hayman and Eklundh, 2003; Mittal and Huttenlocher, 2000). First, image registration is performed by finding corresponding features using a tracker, such as KLT (Lucas et al., 1981). Second, a mosaic is created using projective transformations. Finally, a Gaussian Mixture Model (GMM) (Stauffer and Grimson, 1999) calculates the value of each pixel of the background panorama. The main limitation is that, when the camera is translated, the one-to-one mapping between a background model and an incoming frame cannot be computed due to parallax induced by the movement of the camera's centre. To overcome this, Jin et al. (2008) proposed a multi-layer panorama approach where each layer corresponds to a homography induced by a different plane. To discover those planes, homographies are iteratively estimated using RANSAC (Fischler and Bolles, 1981). As a result, pixels from an incoming frame are rectified on the panorama based on the homography induced by the plane they lie on. Though that method can deal with depth variation and parallax, it still suffers from errors accumulating during panorama construction.

Approaches based on epipolar geometry address more general motions, including camera translations. Initial motion segmentation is conducted using the fundamental matrix (FM) and the epipolar constraint and, then, it is refined taking advantage of block based appearance models (Kwak et al., 2011). A significant constraint is dependency on accurate initialisation of the appearance model for the first frame. Instead of calculating a per frame FM, Zamalieva et al. (2014a) employed the Temporal FM (Yilmaz and Shah, 2006), where a series of FMs models the epipolar geometry across multiple frames. They are calculated iteratively to identify short-term trajectories or tracklets that maximise the number of inlier tracklets. Thus, tracklets whose points do not lie on the corresponding epipolar lines are associated to foreground. Since all those methods are prone to FM calculation degeneracies (Jebara et al., 1999), a model selection criterion between homography and FM was proposed to deal with variety of camera motions and scene geometries (Zamalieva et al., 2014b). The main drawback of such approach is that foreground pixels the motion of which appears similar to the camera's may be assigned to background planes.

An alternative to camera-based models has been to rely on analysis of pixel motions. Study of long-term trajectories allows estimating a background trajectory subspace where foreground trajectories are considered outliers. Sheikh et al. (2009) calculate iteratively the background trajectory subspace using RANSAC and produce an initial sparse B/F labelling. It is refined based on colour and location cues using Markov Random Fields. Although those methods do not assume any specific camera motion, they still show some limitations. First, they rely upon complete trajectories calculated over a frame window. Second, they fail when orthographic projection is not satisfied. Third, they assume background trajectories occupy the majority of the scene. To overcome them, it has been proposed to group long-term trajectories, even incomplete, based on their affinities using spectral clustering (Brox and Malik, 2010). Since that leads to sparse labelling, that approach was extended to create dense regions by propagating spatially and intra-level the trajectory labels (Ochs and Brox, 2011). However, due to their computational cost, those methods are not suitable for real time applications. To address this, Elqursh and Elgammal (2012) modeled spatial and motion trajectory affinities using a low-dimensional manifold which is updated online. However, since trajectory-based techniques suffer from the fact that long-term trajectories are not always available for all points, it was proposed analysing region trajectories, where motion and area statistics obtained from region trajectories are used as features for learning pedestrian motion (Galasso et al., 2011). However, since that procedure relies on a learning phase, its usage is restricted to detecting moving objects which are present in the training dataset.

As an alternative way for dense pixel analysis, some recent methods rely on optical flow. Narayana et al. (2013) used quantised orientations of flow vectors as depth invariant motion cues. As a consequence, objects with motions different from the predefined translational model are considered moving objects. Then, the number of independently moving objects is estimated automatically using non-parametric clustering. The main drawback is that, since it accounts only for camera translation, it cannot deal with camera rotations. Moreover, it cannot detect moving objects whose flow orientation is similar to the camera's one. Following the same paradigm, Bideau and Learned-Miller (2016) used a combination of optical flow angles and magnitudes to describe motion directions for every pixel. By estimating the global background motion direction, B/F likelihood can be calculated for each moving object. Since results are susceptible to optical flow errors as well as dynamic background (waving trees, waves etc.), Tokmakov et al. (2017) designed a deep learning framework based on optical flow vectors including an object classifier and conditional random fields. Despite those efforts, all these methods still suffer from large depth variability and since they focus on short term motion analysis, parts of a moving object which are initially static in a sequence may not be identified as foreground if they start moving later.

B/F segmentation for freely moving cameras is still a challenging task due in particular to scene variability and the range of possible camera motions often preventing usage of any pre-set camera model or trajectory constraints. As a consequence, a model-free approach only based on evolution may have the potential to handle better segmentation of videos captured by freely moving cameras.

3. Vide-omics paradigm

The proposed genomics-inspired paradigm relies on a one-to-one mapping between nucleotide and pixel values. Thus, a DNA sequence corresponds to an image scanline, both sharing a digital and linear nature. Note that although the 2D structure of images is not exploited by the paradigm, it can be taken advantage of in a post-processing stage. Based on the proposed mapping, a strong analogy can be drawn between aspects of the living cell and a visual surveillance system, as illustrated in Fig. 1. First, they display similar internal organisation: the core data of the cell, its genome, is distributed across a set of chromosomes, whereas images produced by a surveillance system are captured by a set of cameras. Note that genes belonging to the same chromosome are more likely to be inherited together. Second, both types of data evolve with time in a quite gradual manner. Although cell duplication involves a process attempting to make faithful copies of DNA chains, mutations occur, introducing many differences between the original and the new sequences. Likewise, successive images

Table 1
Cell mutation types and possible sources producing equivalent image variations in visual surveillance.

Cell mutation type	**Possible sources producing equivalent image variations**
Substitution	Sensor noise, change in camera gain, change in scene illumination
Insertion	Change in camera angle and/or position revealing previously occluded data, more details in common field of view (zoom in), apparition of a new object in a camera's field of view after motion or zoom out
Deletion	Change in camera angle and/or position introducing new occlusions, less detail in common field of view (zoom out), disappearance of an object from a camera's field of view after motion or zoom in
Duplication	Scene area seen by overlapping cameras
Transposition	Motion of foreground object

generated by a given camera are usually highly similar despite scene variation, sensor noise and changes in camera intrinsic, i.e. focal lens and gain, and extrinsic parameters, i.e. location and rotation. Third, gene duplication is an important genetic process which is believed to play a major role in evolution since the absence of genetic pressure on the copies gives them the opportunity to evolve a novel and/or different function Ohno (1970) - analysis of the human genome has revealed that up to 5% is the results of both intra- and inter-chromosomal recent duplications Bailey et al. (2001). In a video surveillance context, corresponding scanlines captured by cameras with overlapping views can be equated as the sequences of a gene and its duplicates. Table 1 illustrates how mutations that are common in genetics can be equated to the main processes generating variations in a video.

The 'vide-omics' paradigm aims at exploiting those analogies: by adapting the now mature approaches which have been developed to analyse genetics data, videos captured by a surveillance system can be processed in a framework where variability is the expected norm. Benefits of the proposed paradigm are that it does not impose any constraint on the way videos are captured and it does not rely on any motion model. Although 'vide-omics' would allow processing videos produced by a whole visual surveillance system where all cameras are connected through a network of pairwise overlapping fields of view, it is relevant to many single camera scenarios: Table 2 lists analogies between computer vision and bioinformatics tasks, and describes the main components of the associated bioinformatics pipelines.

Note that although exploitation of genomics-based solutions for video analysis is not novel, as Section 2.2 shows, it is the first time that a video processing paradigm has been proposed based on those ideas. For example, although the Video Genome Project offers an elegant genomics-based approach for video comparison (Bronstein et al., 2010), it is dedicated to that application and could not be extended to other related tasks such as single video analysis.

In preliminary work, the relevance of this new paradigm was explored in the relatively simple and constrained application of dense pixel matching (Dieny et al., 2011; Thevenon et al., 2012). Although promising results were produced, those studies also revealed that the most efficient approaches take advantage of scenario constraints. As a consequence, to highlight the value of the proposed general paradigm, a quite challenging task has been selected: foreground extraction from data captured by a freely moving camera.

4. Application to background/foreground segmentation for freely moving cameras

4.1. Vide-omics based segmentation pipeline

The proposed pipeline for background/foreground extraction from videos captured by freely moving cameras (Fig. 2) is based on the vide-omics paradigm: a continuous image sequence can be interpreted as the capture of a single image evolving through time through mutations revealing 'a scene'. Although many of the mutations do not affect the nature of the scene - or background -, e.g. sensor noise, change in camera gain, scene illumination and field of view, others reveal the presence of transient objects - or foreground. Thus, effective detection and analysis of those mutations should allow discriminating between the scene's background and foreground objects.

Since, in bioinformatics, mutation detection and analysis relies on the alignment of genetic sequences using techniques such as the Needleman-Wunsch algorithm (NW) (Needleman and Wunsch, 1970), it is proposed to treat a video, as a set of evolving scanlines here the 2-dimensional nature of images is not exploited. As a consequence, the first step of the pipeline is to establish correspondences between the scanlines of each pair of frames. This step is performed by finding the transformation necessary to align two frames as estimated by the positions of matching salient points. In addition to establishing scanline correspondences in the scene, this procedure allows estimating the amount of overlap between scanline pairs.

Then, the vide-omics module processes each scanline independently to identify pixels associated to transient objects. By concatenating those outputs, foreground is produced for each frame. Finally, since this vide-omics based approach does not take advantage of vertical consistency within a frame, this is addressed during the post-processing stage.

4.2. Proposed methodology

Once scanline correspondence has been established for every frame pair, each scanline of each frame is processed independently to find pixel correspondences among overlapping scanlines and detect outliers which could reveal the presence of foreground objects. First, each scanline is pairwise aligned against each of its corresponding scanlines in all the other frames, (see Section 5.2.1). Those pairwise alignments identify areas where pixels cannot find a match without altering the pixel sequence order. Second, those areas are labelled as either foreground objects or occluded areas by analysing their behaviour across all other scanline alignments. Third, vertical consistency between successive scanlines is exploited by a post-processing step connecting and

Table 2
Analogies between computer vision and bioinformatics tasks. For all of them, the main associated bioinformatics tools and techniques are listed.

Computer vision tasks	**Analogous bioinformatics tasks**	**Associated bioinformatics pipelines**
Dense pixel matching	Identify one-to-one correspondences between sequences to assess if they are evolutionarily related	- Global sequence alignment - evaluation of alignment significance
Content-based image retrieval	Identification of common functional or structural features between evolutionarily unrelated sequences	- Local sequence alignment - evaluation of alignment significance
Foreground extraction	Explore genetic differences between two genomes to identify organism-specific genes and rearrangements	- Global sequence alignment - identification of insertions and deletions (indels) - indel classification as either rearrangement or organism specific
Background reconstruction	Infer most recent common ancestor of a family	- Sequence multiple alignment - creation of a phylogenetic tree - common ancestor reconstructionObject recognition
	Identify biologically meaningful patterns (motifs) to predict a gene/proteins function	- Multiple alignment of motif instances- creation of motif descriptor - sequence scanning - evaluation of hits significance

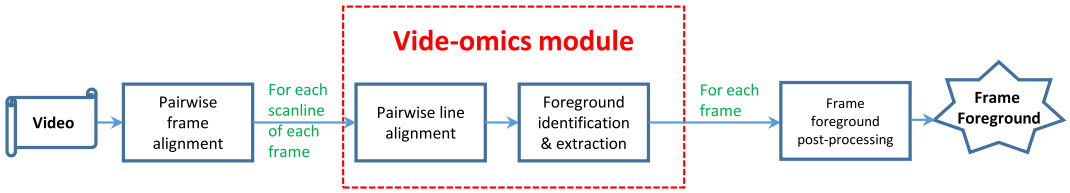

Fig. 2. Description of segmentation pipeline.

merging consistent foreground patches. Next, the methodology is explained in detail.

4.2.1. Pairwise line alignment

In order to find optimal pixel correspondences between pairs of scanlines, it is proposed to use an adaptation of the Needleman–Wunsch algorithm where scanline variations are treated as standard genetic mutations, i.e. pixel insertions, substitutions and deletions. An implementation has already been introduced to address stereo matching between two rectified images (Dieny et al., 2011). It follows closely the NW algorithm which relies on a dynamic programming approach and a scoring function penalising possible mutations. The optimal global alignment of two scanlines is generated in two stages. First, optimal alignments of subsequences starting from the beginnings of the scanlines are calculated and recorded in a table, where each cell contains both the highest score which can be reached by extending a previous partial alignment by one pixel and a link to that previous alignment. The scoring function evaluates if the optimal new alignment should be created by either aligning the next pixel of the first scanline with the next pixel of the second scanline ('match'), or by shifting the unaligned pixels of one of the scanlines by one pixel to model either a pixel insertion or deletion ('gap'). The scoring function penalises poor quality pixel 'matches' with a score based on pixel value difference, whereas the introduction of a 'gap' leads to a fixed penalty. Second, a 'backtracking' phase takes place: the optimal global alignment between the two scanlines is extracted from the table using the optimal alignments of subsequences it has recorded. The NW algorithm is frequently refined by integrating the concept of extended gap (or 'egap') in order to take into account that, in genetics, insertion or deletion of a sequence of n nucleotides is much more frequent that n insertions or deletions of a single nucleotide. As a consequence, adding a 'gap' after an existing 'gap' is less penalised, which encourages 'gaps' to cluster. Since in computer vision, absence of correspondences between pixels captured from overlapping areas usually comes from appearance, disappearance or motion of pixel regions associated to specific objects, the extended gap refinement is also implemented in the scanline alignment algorithm. Further details about this scanline alignment algorithm can be found in Dieny et al. (2011).

4.2.2. Foreground identification and extraction

Pairwise alignments of corresponding scanlines, as shown in Fig. 3 a), highlight pixel regions that cannot be matched in the other scanline. Those regions correspond to moving objects, occluded and/or non-overlapping areas. Foreground identification requires discriminating between these possibilities. The NW algorithm only accounts for three types of mutations, i.e. insertions, substitutions and deletions. Although it is appropriate to represent occluded and non-overlapping areas, which fits well the genetic concept of 'deletion', it has difficulties dealing with the motion of a set of pixels or foreground object between scanlines. As a consequence, it can only represent such pixel motion as both a deletion from one line and an insertion in the other line without recording that the deleted and inserted set of pixels would match each other. Actually, such type of mutation corresponds in genetics to a transposition or a 'jumping gene' discovery by Barbara McClintock (McClintock, 1950) which led to her award of a Nobel Prize in 1983. Since the NW algorithm cannot recognise transpositions, the produced alignments are frequently post-processed to identify 'jumping genes' (Delcher et al., 1999). Following a similar approach, jumping pixel regions in one scanline are identified by searching for matching regions in the other scanline.

The global alignment performed by the NW algorithm highlights regions of a scanline, shown in brown in Fig. 3, which cannot be matched with a region of the other scanline without altering the pixel sequence. As a consequence, those unmatched regions lead to the creation of corresponding gap regions. Those unmatched regions can be classified into 3 distinct categories: (i)Occluded and non-overlapping background areas, ii) foreground objects visible in both scanlines - object motion has some horizontal component[1] and the object is visible in the field of view of the other frame - and (ii) foreground objects visible in only one of the scanlines. On one hand, regions of category (iii) can easily be identified since they have matching regions on both scanlines where both regions are associated to gap regions in the other scanline, e.g. Fa in Fig. 3, Case 1 a). Therefore, the matching of an unmatched region of a given scanline with an unmatched region of another scanline suggests that both regions belong to the same moving object. On the other hand, regions of categories (i) and (iii) share similar properties: they are only visible in one of the two scanlines. As a consequence, such region may not find any good match in the other line, shown in red in Fig. 3, Case 1 (a) and Case 2 (a), and its best match is unlikely to correspond to an unmatched region. Therefore, additional information, i.e. other corresponding scanlines, is required to discriminate between those two categories. On one hand, since occluded and non-overlapping areas belong to the background, their surrounding pixels are consistent across frames. On the other hand, a moving object's neighbourhood tends to vary. It is proposed to exploit that observation in the second part of the foreground identification algorithm. This is performed by comparing the location of the best match of unlabelled regions on other scanlines with the location of the best match of the same unlabelled regions which have been extended to neighbouring pixels. If both locations correspond, one concludes the unlabelled regions belong to the background, see Fig. 3, Case (1 b). If they do not, the unlabelled regions are considered to be foreground regions, see Fig. 3, Case (2 b).

Since pixel region matching is a noisy process and the absence of a region in a scanline leads to an arbitrary best match, decision regarding the belonging of a region to a foreground object cannot be made from a single comparison. As a consequence, each unmatched region is associated to a likelihood of belonging to the foreground. That likelihood is calculated as the number of times the comparison of a scanline of interest to each of its corresponding scanlines led to that region to be labelled as foreground divided by the number of comparisons. The whole algorithm for identifying foreground regions from a given scanline is described by the pseudocode in Algorithm 1. Best matching regions are identified using a sliding window, where, the best match is defined as the pixel block with the lowest sum of square pixel differences (*Best match* function). The extension of an unmatched region of size l is performed by concatenation of its preceding $l/2$ and following $l/2$ pixels from the scanline it belongs to.

[1] Although one cannot assume that foreground objects move horizontally, one can expect that, since objects have usually some vertical homogeneity, matching line fragments can be found between corresponding scanlines for a few frames.

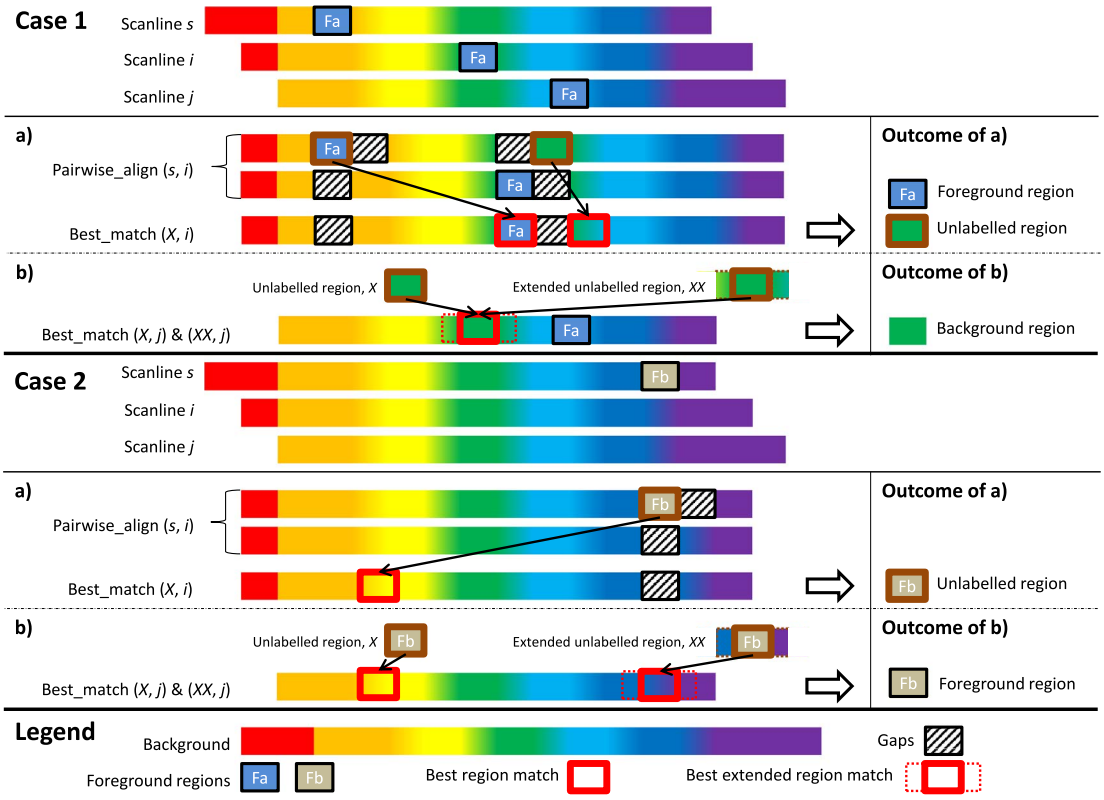

Fig. 3. Illustration of the foreground identification process which relies on a two-stage algorithm: Case 1 depicts a scenario where a foreground object, Fa, is visible in all scanlines which are analysed, occluding various background regions; Case 2 shows a situation where a foreground object, Fb, is only visible in one scanline. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.2.3. Post-processing

Following the processing of all the scanlines of a given frame, a foreground likelihood map is generated. After scaling, it is represented as a greyscale image on which mathematical morphology erosion of size 1 followed by dilation of size 1 - is applied to reduce both background noise and pixels introduced by the algorithm resolution, i.e. 1 pixel resolution. Next, initial foreground segmentation is obtained by, first, extracting all pixels with foreground likelihood above 50% and, then, removing remaining small regions. Finally, in order to take advantage of consistency between adjacent scanlines and connect individual foreground components, it is proposed to use existing foreground patches as seeds to grow consistent foreground regions. The GrowCut algorithm was selected because of its capability to grow regions using sparse foreground and background labelling (Vezhnevets and Konouchine, 2005). Since the proposed method is able to provide highly confident foreground and background regions, GrowCut was employed for further segmentation refinement where regions of initially low confidence can be recovered thank to their high confident surroundings.

5. Experimental results

In this section, experiments are conducted to illustrate the strengths of the vide-omics paradigm. First, the data sets and evaluation framework used to analyse the performance of the proposed algorithm are described. Second, its implementation is detailed. Third, results are presented and discussed.

5.1. Data sets and evaluation framework

Whereas there is a plethora of benchmark datasets from static cameras, there are very few from moving cameras. Moreover, they are usually limited to the rotation motions performed by PTZ cameras. Berkeley Motion Segmentation Dataset (BMS-26) offers a set of 26 videos exhibiting a variety of camera motions and scene geometry complexities which has been widely used for evaluating foreground extraction algorithms (Brox and Malik, 2010). Among them, thirteen representative videos were selected for validation and comparative analysis of the proposed vide-omics approach: people1-2, cars1-10 and marple10. The description of the selected videos can be found in Table 3. On one hand, people1-2 and cars1-10 videos are typical of the output produced by standard PTZ cameras: it involves a small number of objects performing continuous motions in a single scene of low complexity the background of which is unveiled in its entirety. The camera motion consists mainly of rotations with small translations which do not lead to any parallax effect. On the other hand, the marple10 video is a much more challenging video due to additional camera's translation, inducing a large parallax effect, and the complex geometry of the scenes. As a consequence, it has proved particularly challenging for algorithms relying on particular camera and/or scene models and should allow highlighting the value of the model-free vide-omics pipeline. The marple10 video includes three moving and interactive objects, i.e. 'Miss Marple', a man and a cart, where 'Miss Marple' and the man display continuous motions.

Those videos are provided with ground-truth frames with segmented foreground as indicated in Table 3. The ground truth segmentations provided with Marple10 was originally designed for a segmentation task involving a wall and the three moving objects. As a consequence, they are suitable for motion segmentation if the wall is removed from the relevant ground-truth frames, i.e. 1, 10, 20, 30, 40, 50, 100, 150, 200, 250 and 300. Unfortunately, many authors did not perform that adjustment, which makes performance comparison with their approaches difficult.

Performance of the proposed vide-omics pipeline is evaluated against state-of-the art moving object detection methods representing both camera-based models and approaches relying on pixel motion

Data: frame sequence
Result: foreground likelihood for pixels of a given scanline
s: scanline of interest in the frame of index f;
i, j: scanlines;
N: number of frames;
c, t, d: comparison counters initialised to 0;
L(x): foreground likelihood of a pixel region x;
q, r: frame indices, $q \neq f$ and ($r \neq f$ & $r \neq q$) ;
for $q = 1$ **to** N **do**
  i = Corresponding_scanline(s, q); *//scanline corresponding to s in frame q, if null go to q+1*
  $c = c + 1$;
  p = *Pairwise_align*(s,i); *//pairwise alignment between scanlines s and i*
  X = *Unmatched_regions*(s,p); *//regions of s corresponding to gaps according to alignment p*
  for *all* $x \in X$ **do**
    y = Best_match(x, i);
    if $y \in$ *Unmatched_regions(s, p)* **then**
      $L(x) = L(x) + 1$;
    else
      xx = *extended*(x); *//region x is extended to the left and the right by neighbouring pixels*
      $t, d = 0$;
      for $r = 1$ **to** N **do**
        $d = d + 1$;
        j = *Corresponding_scanline*(s, r); *//scanline corresponding to s in frame r, if null go to r+1*
        z = *Best_match*(x, j);
        zz = *Best_match*(xx, j);
        if z *is not a subset of* zz **then**
          $t = t + 1$;
        end
      end
      if $d \neq 0$ **then**
        $L(x) = t/d$;
      end
    end
  end
  for *all x regions* **do**
    $L(x) = L(x)/c$;
  end
end

Algorithm 1. Foreground likelihood quantification for pixels of a given scanline

Table 3
Description of the selected videos.

Video	N. of frames	Frame size	Ground-truth frames
cars1	19	480 × 640	1, 10, 19
cars2	30	480 × 640	1, 10, 20, 30
cars3	19	480 × 640	1, 10, 19
cars4	54	480 × 640	1, 10, 20, 30, 40, 54
cars5	36	480 × 640	1, 10, 20, 36
cars6	30	480 × 640	1, 10, 20, 30
cars7	24	480 × 640	1, 10, 24
cars8	24	480 × 640	1, 10, 24
cars9	60	480 × 640	1, 10, 20, 30, 40, 50, 60
cars10	30	480 × 640	1, 10, 20, 30
people1	40	480 × 640	1, 10, 20, 30, 40
people2	30	480 × 640	1, 10, 20, 30
marple10	460	350 × 450	1, 10, 20, 30, 40, 50, 100, 150, 200, 250, 300, 350, 400, 450, 460

Table 4
Summary of assumptions and limitations of the proposed method and its competitors.

Method	Assumptions/Limitations
Proposed	Scanline-based
HMF	Cannot handle camera translation
PTR	–
FOF	Cannot handle camera rotation
PCM	–
MP-Net+	Rely on a pre-trained object classifier

analysis. Specifically, the proposed method is compared with Probabilistic Causal Model (PCM) (Bideau and Learned-Miller, 2016), a deep learning based framework learning motion patterns (MP-Net + Objectness + CRF using LDOF, referred as MP-Net+ in this paper) (Tokmakov et al., 2017), Point Trajectories to Regions (PTR) Ochs and Brox (2011), Fields of Oriented Flow (FOF) (Narayana et al., 2013) and an implementation of a homography-based method (HMF). HMF creates a panorama of the scene using key frames (Brown and Lowe, 2007): since, every pixel in the panorama is modelled with a median absolute deviation, moving object detection can be achieved by, first, registering a frame of interest to the panorama and then applying background subtraction. Table 4 summarises assumptions and limitations associated to those methods.

Since executables are available for HMF, MP-Net+, PTR and PCM, detailed comparisons could be performed with the proposed method. On the other hand, comparisons with performance of FOF have to rely on published results and, as a consequence, could not be obtained for moving object extraction from the marple10 sequence.

Background/foreground segmentation methods are evaluated according to their ability to distinguish if a pixel belongs to foreground or background class, which is equivalent to a binary classification task. This is achieved through comparison with the ground truth segmentation maps which are associated to the videos of interest. In the context of foreground extraction, true positives (TP) correspond to the number of foreground pixels that are correctly classified as foreground. False positives (FP) are the number of foreground pixels that are classified as background. Conversely, false negatives (FN) are the number of background pixel classified as foreground, whereas true negatives (TN) are the number of background pixels that are classified as background.

Common metrics that are employed for evaluating performance of foreground extraction system are average precision, average recall and the average F1 score. Recall measures the ability of a system to classify correctly foreground pixels penalising the score if background pixels are misclassified as foreground.

$$Recall = \frac{TP}{TP + FN}$$

Precision measures the ability of a system to classify correctly foreground pixels penalising the score if foreground pixels misclassified as background pixels.

$$Precision = \frac{TP}{TP + FP}$$

F1 score combines in a single measure performance in terms of precision and recall. It is often used for comparing overall performance of systems.

$$F1 = \frac{2TP}{2TP + FP + FN}$$

5.2. Implementation details

5.2.1. Frame alignment

Correspondences between the scanlines of two frames are established by, first, matching the salient points identified by KAZE features (Alcantarilla et al., 2012). That feature detector was selected because it outperforms standard methods such as SIFT (Lowe, 1999) and SURF Bay et al. (2006) producing more inliers and a smaller percentage of outliers. This procedure is further refined using Lowe's ratio test with a ratio of 0.7 to only retrieve a set of good quality matches (Lowe, 2004). Second, those matches allow computing a projective transformation between the two frames using RANSAC. Since the proposed algorithm is line based, a line correspondence shift of more than 1 pixel could be critical. As a consequence, the re-projection process is performed iteratively until the maximum number of inliers achieving a re-projection error lower than 1.414, i.e. a maximum error of 1 pixel in both the x and y directions, is identified. Finally, as results produced by the NW algorithm are less noisy when sequences broadly overlap, matching scanlines are further processed so that only overlapping segments remain, see Fig. 4. Here, the overlapping region between two images is defined by the area covered by the matching salient points. Moreover,

Fig. 4. An example of frame alignment and non-overlapping area exclusion. Matching keypoints from frames (a) and (b) define the overlapping segments (c) and (d).

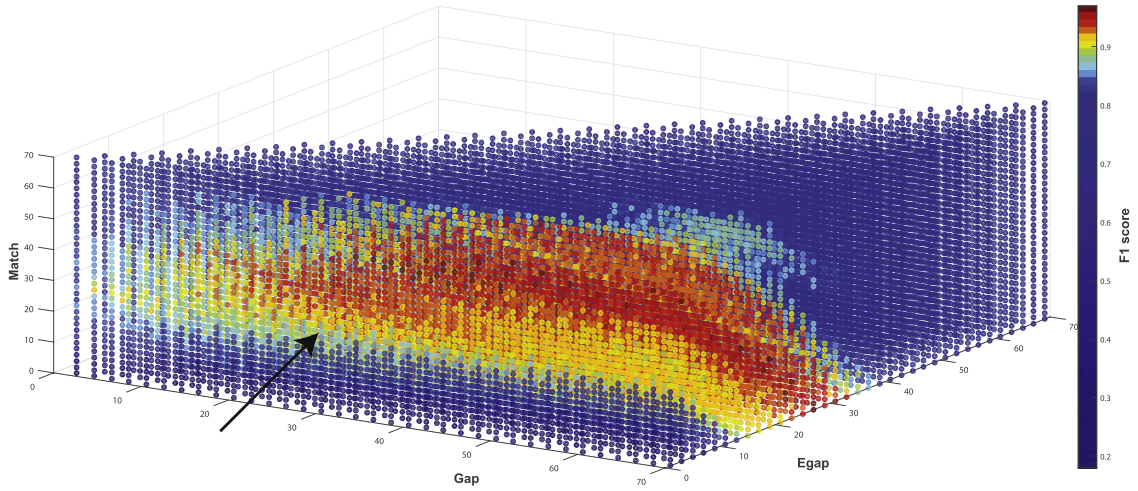

Fig. 5. Exhaustive performance evaluation conducted on the parameter space (*match gap* x *egap*). Colours show F1 scores: dark blue shades show low to average scores, whereas light blue to brown shades show high scores, i.e. $>$ 0.85. Performance of the selected parameters is indicated by the black arrow. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

since alignment significance is affected by sequence size, only scanline pairs the length of which is above 50% of their original size will be pairwise aligned.

5.2.2. Pairwise line alignment

To retrieve pixel correspondences between two scanlines an adaptation of NW algorithm is employed Dieny et al. (2011), where the distance between two pixels d(i,j) - is expressed by the Euclidean distance between their RGB values. That algorithm requires three parameters: *gap, egap* and *match*. While the *gap* and *match* parameters control the balance between introducing a gap and accepting a mismatch, the *egap* parameter promotes the clustering of gaps. While in bioinformatics the selection of those parameters and determination of the optimal substitution matrix have been an active area of research Tripathi et al. (2016); Yamada and Tomii (2014), here the parameter values have been selected experimentally and set to $gap = 30$,$egap = 5$ and $match = 18 - d(i, j)$. The selected parameter configuration ensures that gap introduction is only activated when there is substantial mismatch between two pixel values. In practice, a gap is introduced into the alignment when the distance between two pixels is greater than 48. As a consequence, if the distance between two pixels is greater than 23, an additional gap is inserted. To show that performance is relatively consistent across a wide range of parameter values, exhaustive evaluation was conducted on the parameter space: F1 scores were calculated for the people1 video by processing a single line across the video (scanline 240 on the first frame). Focusing on high F1 scores, i.e. above 0.85 which are represented by light blue to brown colours, Fig. 5 reveals that they are produced by quite a large volume of the parameter space. As a consequence, parameter setting should aim at belonging to that subspace where performance varies quite smoothly. Further analysis also indicates that the *egap* parameter has stronger impact on the results than the *gap* parameter since it has to be selected from a narrower range. This suggests that alignments mainly rely on consecutive gaps the score of which are calculated by $gap + (n - 1)$*$egap$, where n is the number of consecutive gaps. Moreover, the figure shows that, in the high F1 score region, higher *match* leads to lower *egap*: the acceptance of a broader range of mismatches must be compensated by easier creation of gaps.

5.2.3. Foreground identification

Since noisy alignments may lead to generation of a large number of unmatched regions, only continuous unmatched regions of a minimum size are considered as candidates for foreground estimation. Here, a length corresponding to 1% of the width of video frames is chosen. As a consequence, foreground objects of a smaller width can only be recovered during the post-processing step of this methodology. Note that during the foreground estimation process, two regions are established as overlapping if at least 75% of their pixels overlap.

5.2.4. Post-processing

Following the extraction of pixels the foreground likelihood of which is above 50%, unlikely small foreground regions are removed. First, since inaccurate scanline correspondence may lead to isolated foreground lines, those are eliminated if they are only 1-pixel thick. Second, small regions the area of which is lower than the square of 1% of the width of the video frames are also removed. Finally, the resulting foreground objects, f, are used as seeds by the GrowCut algorithm so that the final foreground consists of only a few non-connected foreground components (Vezhnevets and Konouchine, 2005). That approach requires defining three sets containing either background (b), foreground (f) or unlabelled (u) pixels, see Fig. 6. To define the b and u sets, two masks, $m1$ and $m2$, are created by a dilation of the initial foreground by 1% and 2% respectively of the width of the video frames using a disk-shaped structuring element. Whereas the background set is characterised by the pixels which are the farthest away from f in its local neighbourhood, i.e. $b = m2 - m1$, the unlabelled set is defined by the pixels which belong to its most local neighbourhood m1 while not to being part of f, i.e. $u = m1 - f$. Since usage of the GrowCut algorithm as a post-processing step is not standard, performance of the proposed pipeline is provided with and without GrowCut post-processing, i.e. 'Proposed w/o GC'[2].

5.3. Performance evaluation results

Quantitative performance is provided in Table 5 to evaluate the proposed vide-omics pipeline against other state-of-the-art methods which have previously used the people1-2, cars1-10 and marple10 videos. Sequences have been divided in two groups (a) sequences with limited camera motion (people1-2, cars1-10) and (b) a sequence with complex camera motion (marple10)[3]. Regarding foreground extraction from sequences with limited camera motion, results[4] obtained by the proposed method prove to be promising ($\mu_{Prop.} = 0.567$) and outperform

[2] Wall in Marple10 is counted as a foreground object.

[3] This method was evaluated without the last ground truth frame in every sequence since the method cannot process the last frame of a video.

[4] The weighted mean and standard deviation of each group are calculated according to the number of frames in each sequence.

Fig. 6. Illustration of the definition of the three pixels sets used by the GrowCut algorithm. The first column displays the initial foreground set (f), the second and third column show the masks m1 and m2, respectively. The forth column presents the three associated sets: foreground set (f), unlabelled set (u) and background set (b) are coloured in blue, in green and yellow, respectively. The last column shows the final foreground after growth. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 5
Weighted average of F1 scores calculated for each method and group.

	Sequences with limited camera motion (396 frames in total)		Sequence with complex camera motion (460 frames)		
	Weighted Mean	Inter-sequence Std.	Mean	Mean	Intra-frame Std.
Proposed	0.567	0.132	0.576	–	0.155
Proposed w/ o GC	0.505	0.164	0.467	–	0.153
HMF	0.430	0.171	0.303	–	0.196
PTR	0.788	0.167	0.264	–	0.261
PCM	0.776	0.155	0.327	–	0.170
MP-Net + [3]	0.671	0.233	0.404	–	0.296
FOF	0.651	0.144	–	0.580	–

HMF ($\mu_{HMF} = 0.430$). Though PTR, PCM and MP-Net+ display better performance on those sequences ($\mu_{PTR} = 0.788$,$\mu_{PCM} = 0.776$, $\mu_{MP-Net+} = 0.671$), when dealing with the marple10 video they perform quite poorly ($\mu_{PTR} = 0.264$,$\mu_{PCM} = 0.327$,$\mu_{MP-Net+} = 0.404$). This is largely explained by the depth variation present in the scene: the closeness of the wall to the camera generates a strong parallax when the camera translates, resulting into optical flow vectors and long-term trajectories which are very different from those belonging to other background objects. While HMF achieves reasonable results on the first set of sequences, it also performs poorly on the marple10 video ($\mu_{HMF} = 0.303$). This reflects a main limitation of homography which cannot hold when camera translates. Particularly, since the wall occludes another scene which unveils as the camera translates, usage of a single global transformation, i.e. homography, does not allow stitching together frames from different scenes although they share a common plane.

Processing of the more challenging video, i.e. marple10, demonstrates the value of the vide-omics pipeline which shows state-of-the-art performance. Fig. 7, where F1 score is provided for each ground truth frame and camera's motions are annotated, highlights the strength of the new pipeline. Indeed, performance is largely independent from camera motions. On the other hand, in the case of the homographic model-based methodology, HMF, trajectory-based PTR and, to a lesser

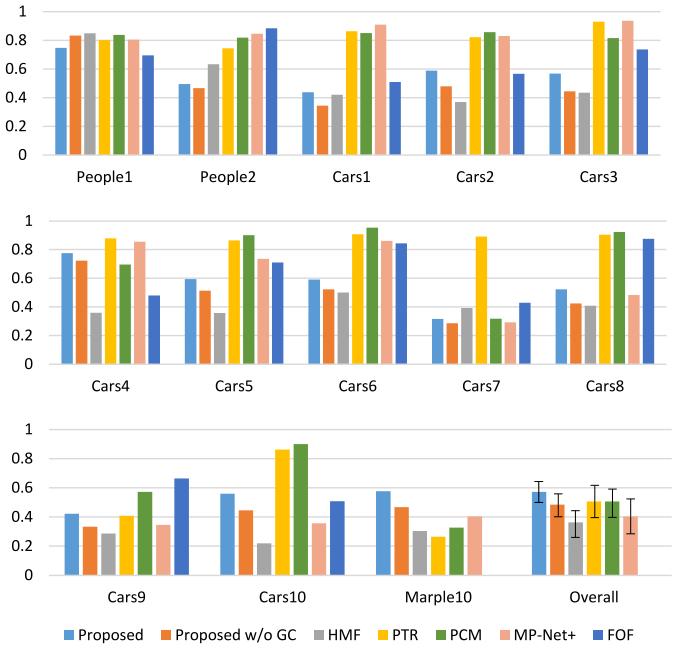

Fig. 8. F1 scores calculated for all sequences and methods.

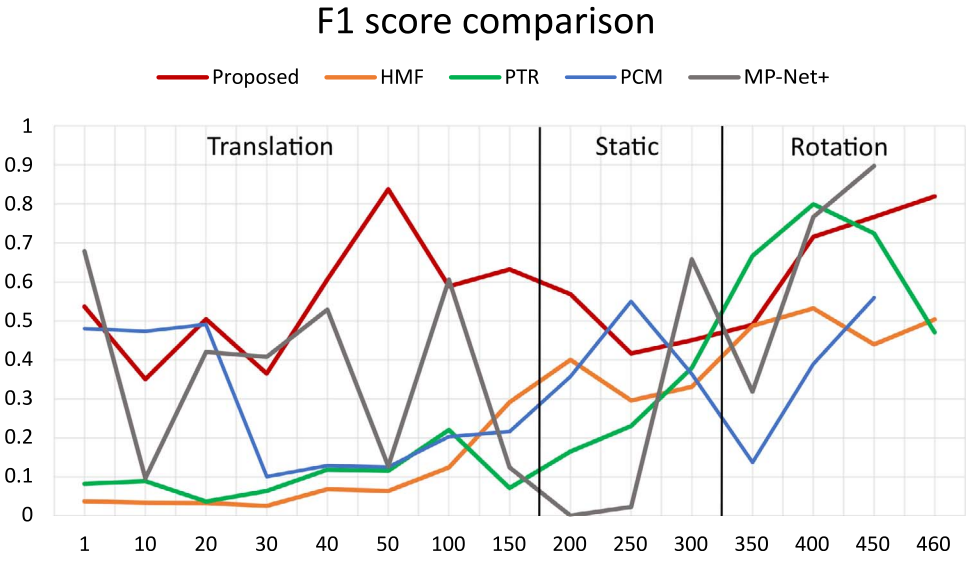

Fig. 7. Foreground extraction evaluation for each frame of the Marple10 video. Note that the type of camera motion is specified for each frame.

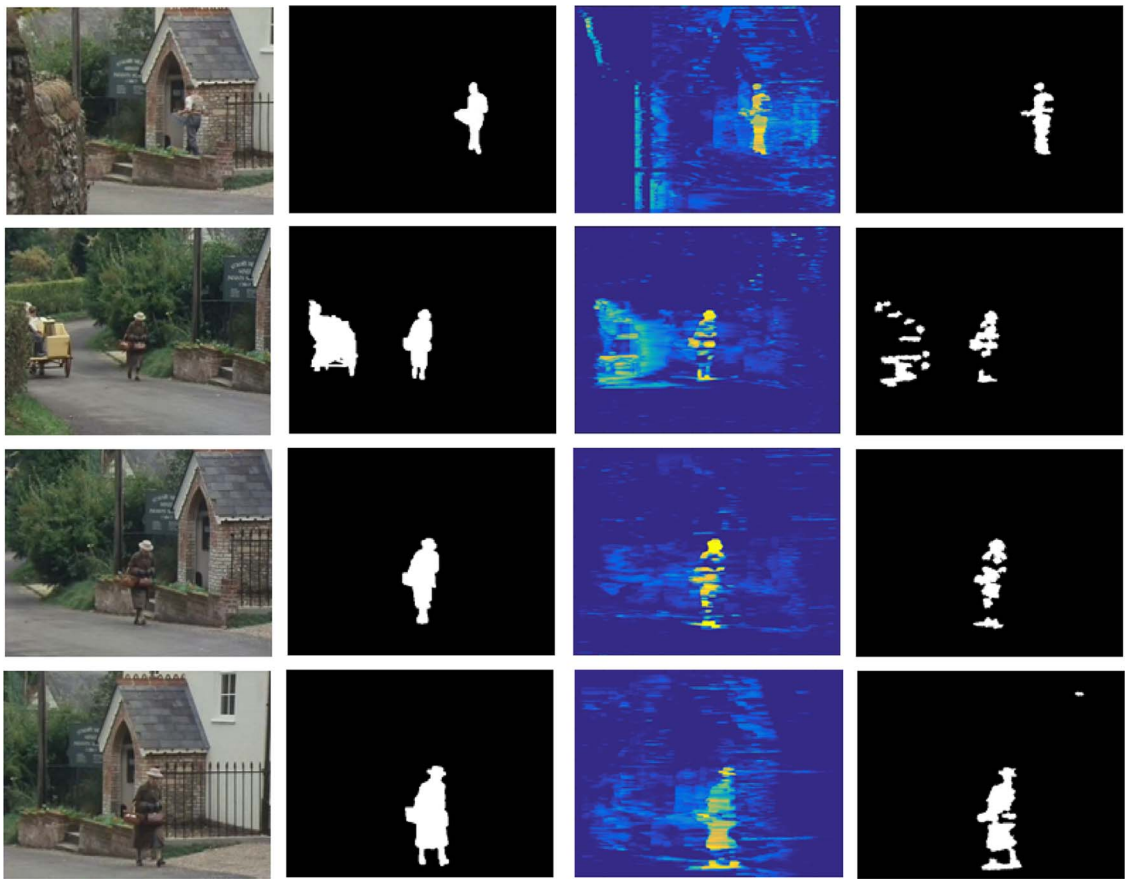

Fig. 9. Examples of foreground extraction using the proposed method for the video Marple10. Frames 50, 300, 400 and 460 are shown in the first column, whereas their associated foreground (ground truth) is presented in the second column. Columns 3 and 4 exhibit the foreground heat map generated after foreground extraction and the detected foreground after post-processing. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

extent, the optical flow based methods PCM and MP-Net+, there is some correlation between the type of camera motion and performance: while those approaches perform well when the camera rotates or is static, they fail to extract adequate foreground when there is camera translation.

Robustness of the proposed method is evaluated, firstly, for each group independently and, secondly, for the two groups combined. For the first group of sequences, the proposed method is the most consistent as shown by a low inter-sequence F1 standard deviation ($\sigma_{Prop.} = 0.132$) compared to ($\sigma_{FOF} = 0.144$), ($\sigma_{PCM} = 0.155$), ($\sigma_{PTR} = 0.167$), ($\sigma_{HMF} = 0.171$) and ($\sigma_{MP-Net+} = 0.233$). For the more complex sequence, the intra-frame standard deviation was calculated to quantify the internal variation of F1 scores. The proposed method ($\sigma_{Prop.} = 0.155$) is also more consistent than any other, i.e. PCM ($\sigma_{PCM} = 0.170$), HMF ($\sigma_{HMF} = 0.196$), PTR ($\sigma_{PTR} = 0.261$) and MP-Net+ ($\sigma_{MP-Net+} = 0.296$). These results reflect the trend of each approach as illustrated in Fig. 7. The proposed method shows a more stable behaviour than other approaches allowing it to deal satisfactorily with a variety of camera motions and scenes. This observation is further supported in the last graph of Fig. 8, where the weighted mean and standard deviation among all sequences are reported. Overall, the proposed method proves to be more consistent across all sequences as demonstrated by inter-sequence standard deviations: ($\sigma_{Prop.} = 0.142$), ($\sigma_{PCM} = 0.165$), ($\sigma_{HMF} = 0.181$), ($\sigma_{PTR} = 0.210$) and ($\sigma_{MP-Net+} = 0.296$). Although the inclusion of GrowCut post-processing significantly impacts F1 performance (up to +23%), it does not affect the main conclusions: the vide-omics pipeline outperforms other approaches in terms of both F1 score and consistency when processing the more complex sequence.

Examples of segmentation results using the vide-omics pipeline are presented in Fig. 9 where extracted foregrounds are compared to initial frames, ground truths and the foreground heat maps generated before post-processing. In those heat maps, foreground likelihood is illustrated using the jet colormap where every pixel value is mapped to a colour using a gradient going from blue (0), to cyan, yellow and red (1).

As expected, this set of experiments has shown that, in constrained scenarios where camera motion is limited, usage of the proposed general paradigm is outperformed by state-of-the-art methods which take advantage of those constraints. However, in the more complex scenario represented by the Marple10 sequence, those methods perform quite poorly, whereas the vide-omics approach achieves significantly better results. Performance in this specific context and the fact that results seem to be much less video-dependent than the other methods provide some evidence of the potential of the vide-omics paradigm.

5.4. Computational complexity

The complexity of the implemented pipeline is dominated by pairwise sequence alignments. This process is performed by an adaptation of the Needleman–Wunsch algorithm whose complexity is $O(nm)$ in both time and memory, where n and m are the lengths of the two sequences. Since the extraction of the foreground associated to a given frame requires the alignment of each scanline of that frame with the corresponding scanlines of a set of k neighbouring frames of identical size (h*w), the time complexity is $O(k.hw^2)$, i.e. $O(hw^2)$, whereas the space complexity is $O(w^2)$ since each scanline is processed independently. As a consequence, the current implementation of the pipeline requires a processing time per frame which is far from being real-time, typically a few minutes using a standard PC with an 8-core processor. Fortunately, the exponential growth of genomics data has conducted the bioinformatics community to design pairwise sequence

alignment techniques with lower computational complexity. First, a modification of the NW algorithm was offered so that optimal alignment could be produced in linear space, while time complexity stayed quadratic Myers and Miller (1988). Addressing computational time, a branch and bound approach has been proposed so that optimal alignment could be produced with a time complexity varying between $O(n + m)$ and $O(nm)$ depending on the similarity between the two sequences, achieving a time gain of 70%–90% for high similarity sequences (> 80%) Chakraborty and Bandyopadhyay (2013). Such implementation would be particularly suitable for the proposed pipeline since neighbouring frames are highly similar in a continuous video. Alternatively, many methods based on heuristics have been suggested to produce alignments in linear time and space, allowing, more than a decade ago, the multiple alignment of 12 entire genomes (including human) in 75 minutes on a PC Brudno et al. (2003). Finally, it has been shown that the NW algorithm is particularly suitable for implementation on hardware platforms (including low cost) (Madeo et al., 2014). As a consequence, the vide-omics pipeline that relies on scanline alignment could be made real-time by using appropriate optimisations, parallel and/or hardware architectures.

6. Conclusion

Based on the principles of genomics, a novel video analysis paradigm, 'vide-omics', has been proposed. Evaluation of its first implementation has provided some evidence of not only its validity, but also its potential. Indeed, using genomics analogies, a background/foreground segmentation pipeline for freely moving cameras has been designed with variability at their core so that performance is constrained by neither camera motions, specific foreground object behaviours nor scene structures. Experimental results showed state-of-the-art performance and robustness when dealing with a challenging video including a variety of camera motions and scene, while remaining competitive in scenes which can be modelled by a specific camera motion model.

One should recognise that initial implementation has limitations which should be overcome to build a system suitable for most real-world applications. First, since scanlines are processed independently from their neighbours, current segmentation does not benefit from vertical spatial coherence. This could be addressed through either an additional post-processing stage which would ensure vertical spatial coherence, combining scanline alignments with 'scancolumn' alignments or a 2D version of the Needleman–Wunsch algorithm which would take into account a pixel's vertical neighbourhood during optimisation of scanline alignment. Second, as discussed, current implementation requires processing times which are far from being real-time. Since usage of heurestics-based alignments has proved particularly efficient in optimising genomics algorithms without altering significantly performance, there is every confidence that such approach would address the high computational complexity of the proposed pipeline.

References

Alcantarilla, P.F., Bartoli, A., Davison, A.J., 2012. Kaze features. Proceedings of European Conference on Computer Vision. pp. 214–227.

Altschul, S.F., et al., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.

Bailey, J.A., Yavor, A.M., Massa, H.F., Trask, B.J., Eichler, E.E., 2001. Segmental duplications: organization and impact within the current human genome project assembly. Genome Res. 11, 1005–1017.

Bay, H., Tuytelaars, T., Van Gool, L., 2006. Surf: speeded up robust features. Proceedings of European Conference on Computer Vision. pp. 404–417.

Bicego, M., Danese, S., Melzi, S., Castellani, U., 2015. A bioinformatics approach to 3d shape matching. Proceedings of European Conference on Computer Vision. pp. 313–325.

Bicego, M., Lovato, P., 2012. 2d shape recognition using biological sequence alignment tools. Proceedings of International Conference on Pattern Recognition. pp. 1359–1362.

Bicego, M., Lovato, P., 2016. A bioinformatics approach to 2D shape classification. Comput. Vis. Image Underst. 145, 59–69.

Bideau, P., Learned-Miller, E., 2016. It's moving! a probabilistic model for causal motion segmentation in moving camera videos. Proceedings of European Conference on Computer Vision. pp. 433–449.

Bouwmans, T., 2011. Recent advanced statistical background modeling for foreground detection - a systematic survey. Recent Pat. Comput. Sci. 4, 147–176.

Bronstein, A.M., Bronstein, M.M., Kimmel, R., 2010. The video genome. Comput. Res. Repos. 2010.

Brooksbank, C., Bergman, M.T., Apweiler, R., Birney, E., Thornton, J., 2014. The european bioinformatics institutes data resources 2014. Nucleic Acids Res. 42, D18–D25.

Brown, M., Lowe, D.G., 2007. Automatic panoramic image stitching using invariant features. Int. J. Comput. Vis. 74, 59–73.

Brox, T., Malik, J., 2010. Object segmentation by long term analysis of point trajectories. Proceedings of European Conference on Computer Vision. pp. 282–295.

Brudno, M., Do, C.B., Cooper, G.M., Kim, M.F., Davydov, E., Green, E.D., Sidow, A., Batzoglou, S., Program, N.C.S., et al., 2003. Lagan and multi-lagan: efficient tools for large-scale multiple alignment of genomic dna. Genome Res. 13, 721–731.

Chakraborty, A., Bandyopadhyay, S., 2013. FOGSAA: fast optimal global sequence alignment algorithm. Sci. Rep. 3, 1746.

Consortium, I.H.G.S., et al., 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–920.

Delcher, A.L., Kasif, S., Fleischmann, R.D., Peterson, J., White, O., Salzberg, S.L., 1999. Alignment of whole genomes. Nucleic Acids Res. 27, 2369–2376.

Dieny, R., Thevenon, J., del Rincon, J.M., Nebel, J.-C., 2011. Bioinformatics inspired algorithm for stereo correspondence. Proceedings of International Conference on Computer Vision Theory and Application. pp. 465–473.

Elqursh, A., Elgammal, A., 2012. Online moving camera background subtraction. Proceedings of European Conference on Computer Vision. pp. 228–241.

Fernandez-Suarez, X.M., Birney, E., 2008. Advanced genomic data mining. PLoS Comput. Biol. 4, 1–7.

Fischler, M.A., Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun ACM 24, 381–395.

Galasso, F., Iwasaki, M., Nobori, K., Cipolla, R., 2011. Spatio-temporal clustering of probabilistic region trajectories. Proceedings of IEEE International Conference on Computer Vision. pp. 1738–1745.

Hayman, E., Eklundh, J.-O., 2003. Statistical background subtraction for a mobile observer. Proceedings of IEEE International Conference on Computer Vision. pp. 67–74.

Jebara, T., Azarbayejani, A., Pentland, A., 1999. 3D structure from 2D motion. IEEE Signal Process Mag. 16, 66–84.

Jin, Y., Tao, L., Di, H., Rao, N.I., Xu, G., 2008. Background modeling from a free-moving camera by multi-layer homography algorithm. Proceedings of IEEE International Conference on Image Processing. pp. 1572–1575.

Kwak, S., Lim, T., Nam, W., Han, B., Han, J.H., 2011. Generalized background subtraction based on hybrid inference by belief propagation and Bayesian filtering. Proceedings of IEEE International Conference on Computer Vision. pp. 2174–2181.

Lovato, P., Milanese, A., Centomo, C., Giorgetti, A., Bicego, M., 2014. S-blosum: Classification of 2D shapes with biological sequence alignment. Proceedings of International Conference on Pattern Recognition. pp. 2335–2340.

Lowe, D.G., 1999. Object recognition from local scale-invariant features. Proceedings of IEEE International Conference on Computer Vision. pp. 1150–1157.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 60, 91–110.

Lucas, B.D., Kanade, T., et al., 1981. An iterative image registration technique with an application to stereo vision. Proceedings of International Joint Conference on Artificial Intelligence. pp. 674–679.

Mackey, A.J., Haystead, T.A., Pearson, W.R., 2002. Getting more from less algorithms for rapid protein identification with multiple short peptide sequences. Mol. Cell. Proteom. 1, 139–147.

Madeo, S., Pelliccia, R., Salvadori, C., del Rincon, J.M., Nebel, J.-C., 2014. An optimized stereo vision implementation for embedded systems: application to RGB and infra-red images. J. Real-Time Image Process. 12, 1–22.

McClintock, B., 1950. The origin and behavior of mutable loci in maize. Proc. Natl. Acad. Sci. 36, 344–355.

McCulloch, S.D., Kunkel, T.A., 2008. The fidelity of dna synthesis by eukaryotic replicative and translesion synthesis polymerases. Cell Res. 18, 148–161.

Medvedev, P., Stanciu, M., Brudno, M., 2009. Computational methods for discovering structural variation with next-generation sequencing. Nat. Methods 6, S13–S20.

Mittal, A., Huttenlocher, D., 2000. Scene modeling for wide area surveillance and image synthesis. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 160–167.

Myers, E.W., Miller, W., 1988. Optimal alignments in linear space. Bioinformatics 4, 11.

Narayana, M., Hanson, A., Learned-Miller, E., 2013. Coherent motion segmentation in moving camera videos using optical flow orientations. Proceedings of IEEE International Conference on Computer Vision. pp. 1577–1584.

Needleman, S.B., Wunsch, C.D., 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48, 443–453.

Notredame, C., Higgins, D.G., Heringa, J., 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. 302, 205–217.

Ochs, P., Brox, T., 2011. Object segmentation in video: a hierarchical variational approach for turning point trajectories into dense regions. Proceedings of IEEE International Conference on Computer Vision. pp. 1583–1590.

Ohno, S., 1970. Evolution by Gene Duplication. Springer-Verlag.

Riedel, D.E., Venkatesh, S., Liu, W., 2008. Recognising online spatial activities using a bioinformatics inspired sequence alignment approach. Pattern Recogn. 41,

3481–3492.
dos Santos-Paulino, A.C., Nebel, J.-C., Florez-Revuelta, F., 2014. Evolutionary algorithm for dense pixel matching in presence of distortions. Proceedings of European Conference on the Applications of Evolutionary Computation. pp. 439–450.
Sebastiani, P., Kohane, I., Ramoni, M., 2003. Machine learning in the genomics era editorial: methods in functional genomics. Mach. Learn. 52, 5–9.
Sheikh, Y., Javed, O., Kanade, T., 2009. Background subtraction for freely moving cameras. Proceedings of IEEE International Conference on Computer Vision. pp. 1219–1225.
Smith, T.F., Waterman, M.S., 1981. Identification of common molecular subsequences. J. Mol. Biol. 147, 195–197.
Stauffer, C., Grimson, W.E.L., 1999. Adaptive background mixture models for real-time tracking. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 246–252.
Thevenon, J., del Rincon, J.M., Dieny, R., Nebel, J.-C., 2012. Dense pixel matching between unrectified and distorted images using dynamic programming. Proceedings of International Conference on Computer Vision Theory and Applications. pp. 216–224.
Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.
Tokmakov, P., Alahari, K., Schmid, C., 2017. Learning Motion Patterns in Videos. Proceedings of IEEE Conference on Computer Vision & Pattern Recognition, CVPR 2017.
Toyama, K., Krumm, J., Brumitt, B., Meyers, B., 1999. Wallflower: Principles and practice of background maintenance. Proceedings of IEEE International Conference on Computer Vision. pp. 255–261.
Tripathi, A., Gupta, K., Khare, S., Jain, P.C., Patel, S., Kumar, P., Pulianmackal, A.J., Aghera, N., Varadarajan, R., 2016. Molecular determinants of mutant phenotypes, inferred from saturation mutagenesis data. Mol. Biol. Evol. 33 (11), 2960–2975.
Vezhnevets, V., Konouchine, V., 2005. GrowCut: Interactive multi-label ND image segmentation by cellular automata. Proceedings of Graphicon. pp. 150–156.
Yamada, K., Tomii, K., 2014. Revisiting amino acid substitution matrices for identifying distantly related proteins. Bioinformatics 30 (3), 317–325.
Yilmaz, A., Shah, M., 2006. Matching actions in presence of camera motion. Comput. Vis. Image Underst. 104, 221–231.
Zamalieva, D., Yilmaz, A., Davis, J.W., 2014. Exploiting temporal geometry for moving camera background subtraction. Proceedings of International Conference on Pattern Recognition. pp. 1200–1205.
Zamalieva, D., Yilmaz, A., Davis, J.W., 2014. A multi-transformational model for background subtraction with moving cameras. Proceedings of European Conference on Computer Vision. pp. 803–817.
Zhang, X., 2007. Pattern recognition in mining high-throughput genomics/proteomics data: the new challenges in old questions. Proceedings of International Conference on Computing Theory and Applications. pp. 242–244.