SENSAREA, a general public video editing application

Pascal Bertolino

SENSAREA, A GENERAL PUBLIC VIDEO EDITING APPLICATION

Pascal Bertolino

GIPSA-lab, Grenoble Alpes University

ABSTRACT

In this demonstration, we present an advanced prototype of a novel general public software application that provides the user with a set of interactive tools to select and accurately track multiple objects in a video. The originality of the proposed software is that it doesn't impose a rigid *modus operandi* and that automatic and manual tools can be used at any moment for any object. Moreover, it is the first time that powerful video object segmentation tools are integrated in a friendly, industrial and non commercial application dedicated to accurate object tracking. With our software, special effects can be applied to the tracked objects and saved to a video file, and the object masks can also be exported for applications that need ground truth data or that want to improve the user experience with clickable videos.

Index Terms— Video editing, video object segmentation, object tracking, ground truth annotation, special effects

1. INTRODUCTION AND MOTIVATIONS

We present our application that is half way between professional video editing software and video annotation tools used in the research domain of video processing and analysis.

Many offline video processing applications need that specific characters or objects be accurately cut out in each frame: in cinema post-production, the needs are considerable, with for examples special effects, movie colorization or mono to stereo conversion. But accurately delineating moving or deformable objects in hundreds or thousands of frame is a tedious task. To that end, a couple of commercial software embed complex and powerful tools (like planar tracking [1]) that help the user to perform semi-automated tracking. Among them, Adobe After Effects and Premiere, Apple Motion and Final Cut, Video Deluxe by Magix and Sony Vegas. Some companies have developed specific tracking addons for these software, like Imagineer Systems with Mocha. However, since these software are very powerful and can perform any kind of 2D or 3D processes, they are very complex to handle and quite expensive. In short, they are not suited for the average user.

In the research field, video object mask extraction for ground truth annotation becomes essential to build database knowledge for machine learning and algorithms evaluation. So, in parallel to the development of the above cited commercial applications, since the end of nineties, researchers have developed software tools such as Videoprep and VideoClic [2] or Qimera [3], to segment or track the content of videos. Some other handy interactive (and sometimes online) software solutions were then developed, mainly based on bounding box interpolation or tracking, to collect large databases of objects for video annotation purpose: VIPER-GT [4], VATIC [5], LabelMe [6], ANVIL [7], SVAT [8].

In [9], we proposed a first version of Sensarea that was dedicated to accurate video object tracking only. The version that we present here (figure 1) is functionally and technically enriched with particularly (1) the possibility to apply effects to the tracked objects, (2) a novel video object tracking algorithm and (3) a key frame based morphing. With this new release, we want to gather both the power of several tracking algorithms and the necessary set of tools to possibly edit/correct the tracking results in a very intuitive application, without any *modus operandi*. To our best knowledge, it is currently the only available software of this kind. Since it is a good tradeoff between simplicity and efficiency, it is suited to anyone who wants to enrich or extract the content of a video in an object based manner. A short demo of Sensarea is available on the web[1] as well as the Windows version of the application[2].

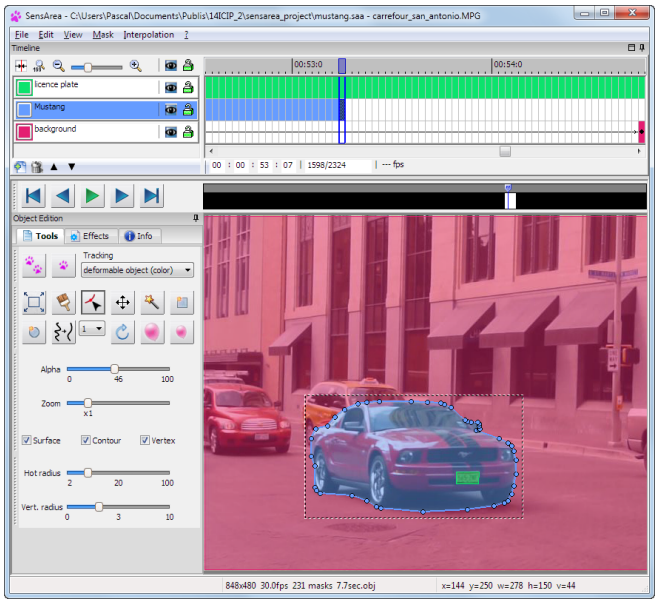

Fig. 1. The Sensarea interface. Top: the layers and the time line. Left: the drawing and tracking tools. Bottom right: the canvas where the user interacts with the frame content

2. SCIENTIFIC AND TECHNICAL DESCRIPTION

Among all the tools available in Sensarea, the most powerful and time saving ones are the tracking tools. At the moment, two different algorithms are available: the first one is based on the irregular graph

[1] www.youtube.com/watch?feature=_embedded&v=5kcBU0haACI

[2] www.gipsa-lab.grenoble-inp.fr/~pascal.bertolino/software.html

pyramid and is described in [10]. The second one is a novel graph-cut based algorithm that works as follows: its input is an object mask at time t noted $M(t)$ provided by the user (using the magic wand or any other drawing tool). In $M(t)$, the pixels are labelled either object or background. Image $I(t)$ is projected to image $I(t+1)$ using a block matching algorithm. The set of corresponding motion vectors are then used to project $M(t)$ to $\hat{M}(t+1)$, a temporary and rough mask of frame $t+1$. Since in $\hat{M}(t+1)$ some pixels have no label and since block matching cannot cope with local deformation or rotation, $\hat{M}(t+1)$ must be refined to better fit with the content of $I(t+1)$. To that purpose, a narrow strip of pixels at the junction of object and background has its labels removed in $\hat{M}(t+1)$. All the non labelled regions are dilated to reconsider non reliable pixel labels. Then, the $\hat{M}(t+1)$ object and background pixels are used as hard constraints in a graph cut process [11] that classically allocates non labelled pixels to the object or background to get $M(t+1)$.

This algorithm provides a good localization of the object borders and it is tuned to behaves correctly in most of the cases without any parameter modification from the user. For big masks, it runs faster than the previous algorithm, and it is fast enough to process several frames in a seconds. Actually, it has no needs to be too fast since the user needs time to visually check on the fly the correctness of the tracking to possibly stop it and apply some manual corrections.

3. IMPLEMENTATION AND USE

Unlike numerous cutting edge algorithm prototypes that cannot be used in a production context (because of a lack of environment tools for initialization, correction and data management storage), our video processing algorithms are embedded in a full functional graphic application that has been tailored to run fast with a lot of data in memory.

3.1. The application components

Sensarea has been developed in C/C++. All the image processing tools are written in C and all the external libraries are cross-platform: the graphic interface is coded in C++ using the wxWidgets library [12]. Some basic processes are done using OpenCV. The database engine that stores all the results is SQLite, a C library. This latter also allows to manage a powerful undo/redo functionality. The video decoding and coding is done using several FFMPEG libraries. At the moment, the application runs under Windows but it could be compiled under Linux or MacOS.

In order to allow a realtime access to any frame and its content, an optimized data structure has been developed and multithreading allows efficient concurrent tasks such as decoding frames, running the tracking, applying the effects and displaying the results on the fly while taking into account the user interactions.

3.2. User interaction

The user starts by opening a video file (all common file formats are supported as well as numbered image sequences) or drag and drop the file in the application window. A default empty layer is ready to receive the first mask provided by the user. This mask can be made using any combination of the following tools: brush, polygon vertices, magic wand, eraser, rectangle or ellipse. Some tools access the masks as raster bitmaps (brush, eraser, mask erosion, delation or simplification) while others access them as vectorized shapes (polygon, rectangle, ellipse). Then a click to the track button tracks the corresponding mask in the successive frames showing on the fly the results. Another click on the same button stops the tracking. Then the user can browse the video to go back and forth, to modify or correct a mask and relaunch the tracking from a corrected mask. If another object needs to be tracked in the same frames, a single click creates a new layer, independent from all the other layers. Any mask of any layer can be edited at any moment. A powerful, flexible zoom allows pixel accuracy when needed.

When the object motion and deformation is rather linear between two frames, it can be advantageous to generate the intermediate masks using our own interpolation algorithm. Inserting interpolated masks between two existing masks (figure 2, top) is obtained by right clicking in the time line between the corresponding (key)frames. Thus any modification of a keyframe mask leads to a real time computation of the interpolated masks (figure 2, bottom).

At any time, it is possible with a click to apply one or several effects to the masks of a layer and to see the results in real time. Effects can be turned off or changed since the original video itself is not modified (figure 3).

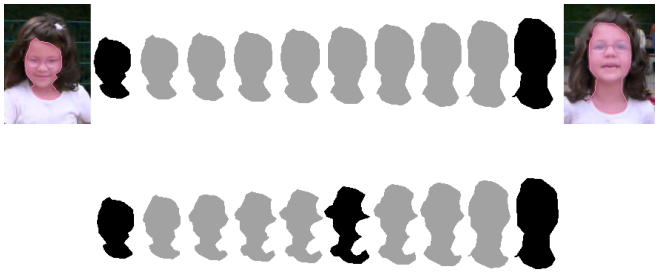

Fig. 2. Example of fully automatic morphing. Top: in a first time, morphing is applied between the 2 black masks. Bottom: Then, one of the interpolated mask is arbitrary modified (the black one) and all the interpolated masks are updated accordingly

Fig. 3. Example of different effects applied to the background and the car, on a given frame with the same car and background masks (top left is the original frame, bottom right shows the zoom effect)

3.3. About the demo

Several short videos will be provided and the user will be able to use both the tracking and the drawing tools to track one or several objects of interest on several frames. Then he/she will choose and tune some effects (figure 4) to apply to these objects. At last, it will

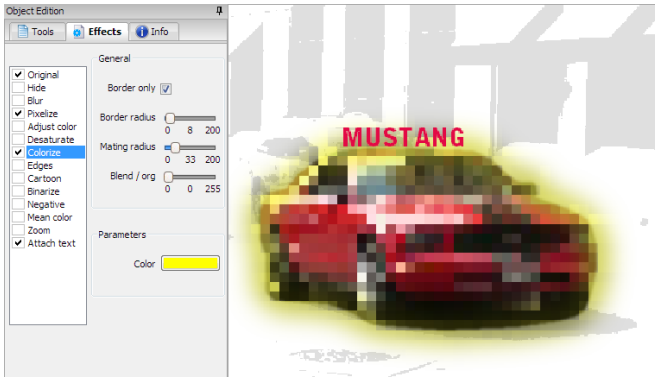

Fig. 4. The effect tab is used to easily apply one or several effects to each layer. Each effect can be finely tuned. In this example, the car keeps its original colors while being pixelized. Its border is colorized in yellow and a text is attached to it

be possible to export the edited video or the object masks and to play the video in our special HTML5 clickable video player.

4. CONCLUSIONS AND FUTURE DEVELOPMENTS

We have presented an application that offers to the average user a novel and flexible way to perform accurate object tracking and special effects. The main strengths of our solution are simplicity, interoperability and interactivity. In a future major release, we will embed a new powerful tracking algorithm [13]. We also want to develop a pluggin manager so that anyone can add his/her own tracking algorithms. We plan to process RGBD videos since there is not yet any practical solution to take benefit of both colour and depth data to improve the segmentation process.

5. REFERENCES

[1] G. Simon, A. Fitzgibbon, and A. Zisserman, "Markerless tracking using planar structures in the scene," in *Proc. International Symposium on Augmented Reality*, Oct. 2000, pp. 120–128.

[2] S. Benayoun, H. Bernard, P. Bertolino, P. Bouthemy, M. Gelgon, R. Mohr, C. Schmid, and F. Spindler, "Structuring video documents for advanced interfaces," in *ACM Multimedia*, Bristol, UK, september 1998.

[3] N. O'Connor, T. Adamek, S. Sav, N. Murphy, and S. Marlow, "Qimera: A software platform for video object segmentation and tracking," in *In Proc. Workshop on Image Analysis For Multimedia Interactive Services*, 2003, pp. 204–209.

[4] D. Doerman and D. Mihalcik, "Tools and techniques for video performance evaluation," in *International Conference on Pattern Recognition*, 2000, vol. 4, pp. 167–170.

[5] Carl Vondrick, Donald Patterson, and Deva Ramanan, "Efficiently scaling up crowdsourced video annotation," *International Journal of Computer Vision*, pp. 1–21, 2012, 10.1007/s11263-012-0564-1.

[6] B. C. Russell, A. Torralba, K. P. Murphy, and W.T. Freeman, "Labelme: A database and web-based tool for image annotation," *International Journal of Computer Vision*, vol. 77, no. 1-3, pp. 157–173, 2008.

[7] M. Kipp, *Multimedia Annotation, Querying and Analysis in ANVIL*, chapter 19, Multimedia Information Extraction, Wiley, IEEE Computer Society Press, 2012.

[8] P. Schallauer, O. Sandra, and H. Neuschmied, "Efficient semantic video annotation by object and shot re-detection," in *2nd International Conference on Semantic and Digital Media Technologies (SAMT)*, Koblenz, Germany, 2008.

[9] Pascal Bertolino, "Sensarea: An authoring tool to create accurate clickable videos," in *10th International Workshop on Content-Based Multimedia Indexing*, Annecy, France, june 2012.

[10] Guillaume Foret and Pascal Bertolino, "Label prediction and local segmentation for accurate video object tracking," in *SPIE Visual Communications and Image Processing*, Lugano, Switzerland, 8-11 July 2003.

[11] Y.Y. Boykov and M.P. Jolly, "Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images," *International Conference on Computer Vision*, vol. 1, pp. 105–112, July 2001.

[12] Julian Smart, Kevin Hock, and Stefan Csomor, *Cross-Platform GUI Programming with wxWidgets (Bruce Perens Open Source)*, Prentice Hall PTR, Upper Saddle River, NJ, USA, 2005.

[13] J. Gallego and P. Bertolino, "Foreground object segmentation for moving camera sequences based on foreground-background probabilistic models and prior probability maps," in *ICIP*, Paris, France, October 2014, To appear.