

Tracking of video objects using a backward projection technique

Stéphane Pateux

IRISA/INRIA, Temics Project
Campus Universitaire de Beaulieu
35042 Rennes Cedex, FRANCE

ABSTRACT

In this paper, we present a technique for tracking video objects in a sequence. The proposed technique is based on a backward projection technique. Since classical backward technique can be disturbed by occlusions and potential errors in spatial segmentation or motion estimation, we propose an extension to the backward projection technique in order to cope with these problems. Results obtained show the relevance of the proposed approach for various kind of tracked objects that can either be rigid or non-rigid.

Keywords: video segmentation, object tracking, backward projection

1. INTRODUCTION

With the development of object-based treatments (such as object-based coding used in MPEG4), the needs for object extraction in video sequence is becoming more and more important. In studio, object extraction could be performed by using chroma-keying, or using a knowledge on a static background¹. However for natural scenes, this technique can't be generally used since background is not always perfectly known.

Lots of work are then done in order to help extract objects from natural scenes. Techniques proposed can be generally described by a two-step algorithm; detection of the objects followed by a tracking of these objects. In this paper we will focus on the tracking technique. We will then assume that an initial segmentation into objects of an image is known, and we will look after a tracking over time of this initial segmentation*. In order to perform this tracking, most of the existing works use a forward technique. This technique consists in projecting the segmentation map known at time t onto the frame $t + 1$ (or further away in time) according to motion information. However this projection step is not generally accurate due to the simplicity of the projection model and to the apparition of occlusion areas. Therefore a post-treatment is generally needed in order to adjust the boundaries of the objects and to treat the occlusion areas. For this purpose, many techniques have been proposed such as morphological watershed², active contours³, or even clustering techniques⁴. In all these techniques this adjustment step aims at ensuring a good localization of the boundaries that is near spatial discontinuity.

Recently, backward tracking technique has been proposed as an alternative method in order to perform object tracking⁵⁻⁷. The main advantage of this technique is the use of a spatial segmentation on each frame in order to ensure a good localization of boundaries (thus no adjustment step is needed), and the ability to deal with non-rigid objects where forward tracking techniques are limited.

In this paper, we propose an extension of the previously proposed backward tracking technique in order to improve it in presence of occlusions or bias in spatial segmentation and motion estimation. In section 2 we present the principle of the backward tracking technique. In section 3 a discussion on the defaults of the backward tracking technique is done and modifications are proposed in order to improve the tracking algorithm. Finally section 4 shows some results and then conclusion and perspectives are raised in section 5.

Further author information:

E-mail: stephane.pateux@irisa.fr

URL: <http://www.irisa.fr/temics/>

*In our study, initial segmentation is manually obtained with the use of an interactive graphic user interface.

2. BACKWARD TRACKING OF OBJECTS

2.1. Principle of the backward tracking

In backward tracking techniques⁵⁻⁷, object masks on a frame are defined with the use of a classification performed over a spatial segmentation. This classification is done using a backward projection onto a reference frame according to the motion of this region (see figure 1); the label of each region is assigned to the label of the object mask it projects mostly on.

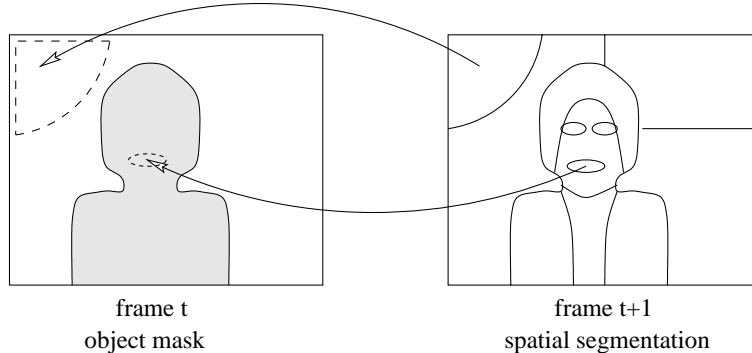


Figure 1. Backward projection technique. Each region of frame $t + 1$ is backward projected onto frame t where the object mask is known; it is then considered to be part of the object it backward projects on.

Compared to forward based technique, this technique presents many advantages:

- **good localization of the boundaries:** since objects are defined as a set of spatial regions, boundaries of objects should be well located on spatial discontinuities. Therefore in previous works on backward tracking technique, no adjustment step is used⁵⁻⁷.
- **tracking of rigid or non-rigid objects:** since an object is defined as a set of regions that could have different motion, these objects could be non-rigid. The use of this transient over-segmentation permits to model more accurately the deformations of the objects. See for example the case of the mask of a person, its deformation can't be modeled with a global motion, but can be defined by a set of moving parts (i.e. various parts of the body; arms, legs, ...). This over-segmentation of objects has also been used in forward technique in order to improve the prediction step⁸.
- **occlusion areas:** with the help of the spatial homogeneity of regions, occlusion areas are affected accordingly to the spatial region they belong to. Small uncovering an covering areas are then well affected. In forward techniques, covering areas can be easily treated, while uncovering areas generally are treated in the adjustment step.
- **simplicity:** the tracking algorithm is rather simple to implement and does not have a high complexity. Main task are just spatial segmentation and motion estimation that are not too complex.

2.2. Spatial segmentation

Spatial segmentation is needed in order to perform the backward tracking technique. This spatial segmentation should give regions with good localization of the boundaries and provides an over-segmentation in order to be able to define objects as a set of regions. Classically segmentation in regions with gray level homogeneity are used: morphological watershed^{5,7}, region growing according to gray levels homogeneity⁶.

However, since motion estimation has to be performed on spatial regions, regions shouldn't be too homogeneous in order to be able to have a good motion estimation. In this paper, we propose to use the spatial segmentation presented in⁹; this segmentation is based on the MDL (Minimum Description Length) formalism and is well adapted for the segmentation of color images. Within this segmentation tool, a parameter [†] permits to define the level of

[†]this parameter can be interpreted as a penalization cost for the presence of a region.

segmentation from an over-segmented results to a segmentation with only a few regions (see figure 2 for an example of results). Furthermore this parameter is quite stable and the same value can be used on many sequences.

Typically for this purpose, the penalization cost is set to a value around 1000. Using lower values leads to segmentation with too small regions leading to difficult motion estimation and misclassification. Using higher values leads to region that could overlapped objects (see right shoulder of foreman for $DL(\theta) = 3200$ on figure 2).



Figure 2. Example of spatial segmentation with different levels of details. Segmentation are obtained with different values of the penalizing cost ($DL(\theta)$ in⁹).

2.3. Motion compensation

In order to project spatial regions at time $t + 1$ on frame at time t , affine motion model is used. This affine motion model can handle translation, rotation, zoom and stretching, which makes it more suitable than a simple translation model such as in⁶.

Motion estimation is based on the minimization of the prediction error in a multi-resolution fashion as done in³. Relaxation on each region is also performed with estimated motion of neighbor regions in order to get a more accurate and stable motion field.

3. MODIFIED BACKWARD TRACKING TECHNIQUE

3.1. Artifacts introduced by the backward tracking technique

However, even if⁵⁻⁷ show promising results for the backward tracking technique, this technique can also encounter problems. For example, if the spatial segmentation is not good enough or when motion is not accurately estimated[‡], there could have a spatial region which is backward projected over two different objects (see figure 3). This region will then be affected to the object it mostly backward projects on.

[‡]This is especially the case for small regions.

For example, in the case of the foreman sequence, let us assume that the object segmentation is composed of the man and the background. Then if spatial segmentation is the one of figure 2 with $DL(\theta) = 3200$, the spatial region on the bottom right that contains background parts and the man shoulder will be affected to the background. In this case, the man shoulder will be lost.

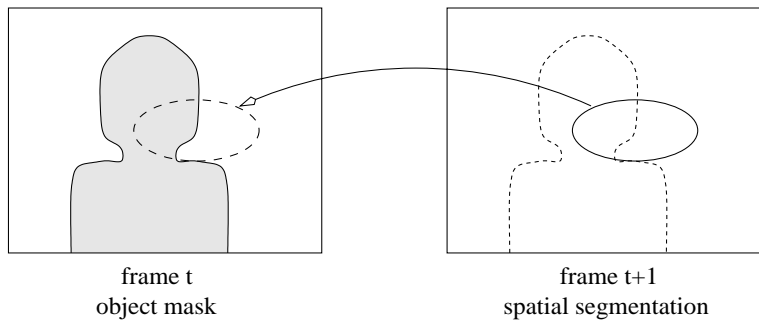


Figure 3. Backward projection of a region over two different objects.

3.2. Origin of the artifacts and possible handling

In fact when looking more closely at the case where a region is backward projected onto several objects, there are many reasons for this kind of misclassification to happen:

1. **default in the spatial segmentation** (especially if regions are too coarse). A spatial region is part of several objects and then when is backward projected could be affected to any of these objects.
2. **default in the motion estimation** (especially if regions are too small). This default can occur in presence of occlusion on the spatial region, thus disturbing the motion estimation (even when using robust motion estimator).
3. **occlusion areas** (see figure 4). Spatial regions can be backward projected on any of the objects that produce the occlusion.
4. **object with changing aspect**. In this case no perfect projection can be performed. However pixels that backward project on the wrong object are a minority, and classification of the region to the object it mostly projects on works well.

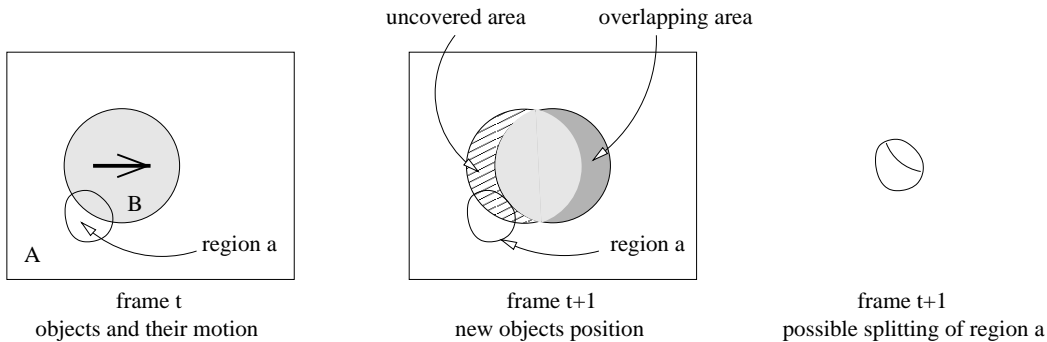


Figure 4. Occlusion areas and backward projection. When a spatial region in frame t contains uncovered areas, this uncovered areas can't be matched in the previous frame. On the other hand, regions containing covered areas can be projected on object A or object B depending on the relative depth between objects.

For example, on figure 4, region a , when ideally backward projected, will lay on objects A and B. This is due to the fact that uncovered areas shouldn't be backward projected since there is no correspondent areas in previous frame.

Intuitively the first three cases could be treated by making a splitting of the concerned spatial regions. In the first case, spatial regions should be split accordingly with the object definition. In the second case, the splitting could help estimate motion by minimizing prediction error only on pixels that are matched. In the third case, the splitting will be used in order to identify uncovering areas that could lead to misclassification.

3.3. Modified backward tracking algorithm

From the previous remarks, we now propose to improve backward tracking technique by the use of a splitting technique of spatial regions whenever it is needed.

First the splitting of regions is based on an association of a region with an object with a given motion (motion of the concerned object). In this case, when region is backward projected, this region can be split into two areas: areas projecting on the selected object, and areas projecting elsewhere (see scheme on the right of figure 4). This basic splitting technique allows also to improve motion estimation by making motion estimation between a region and the texture of an object instead of making motion estimation between a region and the texture of an image. In the first case, occlusion areas are identified and not taken into account in the motion estimation whereas in the second case occlusions are not detected and lead to biased motion estimation.

In the case where a spatial region belongs to many objects, several backward projections are performed with the motion relative to each object. Figure 5 shows such an example and the corresponding splitting. Several areas can be defined:

- **areas that are matched to only one object.** These areas can then be directly classified as being part of this object.
- **areas that are matched to several objects.** These areas correspond to objects occluding themselves.
- **areas that are not matched.** These areas correspond to uncovering areas.

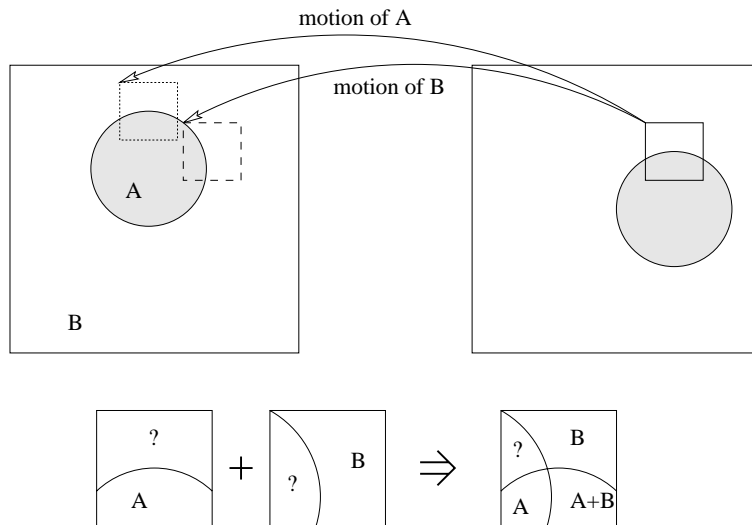


Figure 5. Backward projection of a region using several motion propositions. At the bottom is shown affectation relative to each tested motion, and merging of affectations.

Occluding areas can be treated by affecting them to the object which gives the best prediction. This affectation can be made globally or pixel by pixel with eventually markov random fields technique for consistency. In this paper we use a simple direct pixel affectation to the object with the best prediction for this pixel.

In order to treat uncovering areas, we choose to use the spatial coherence. That is uncovering areas are affected to the object on which the spatial region backward projects on mostly.

3.4. New backward tracking algorithm scheme

As been seen in the previous section, modified backward tracking needs to have matches between regions and objects. To this purpose, we propose the algorithm scheme presented on figure 6.

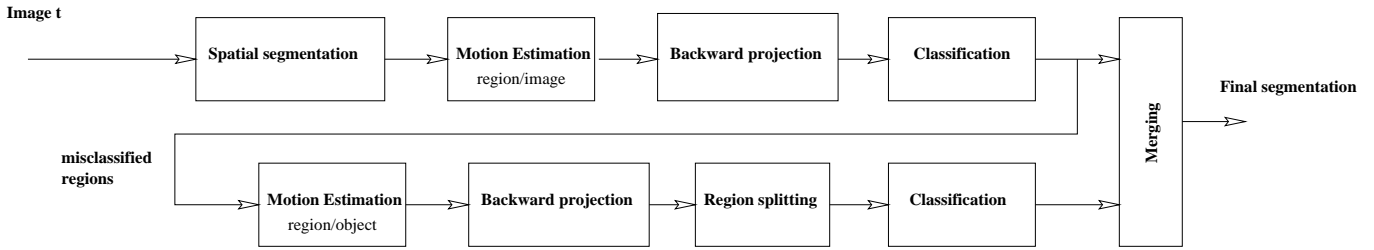


Figure 6. Modified backward tracking scheme.

In a first step, motion estimation relative to the previous image is performed for each region, and backward projection is performed for each region. Regions that projects on only one object are directly classified, other are post-processed in a second step.

The aim of the second step is to split regions as presented in figure 5. To this purpose motion estimation is first refined. Motion is estimated for each possible matching couple of $(region, object)$ as presented in the previous section. The set of possible matching couple is obtained within the first step of the algorithm thanks to the backward projection; for each object on which a region is partly projected a new couple $(region, object)$ is defined. Motion estimation is performed with a Gauss-Seidel minimization of the prediction error for matched pixels; initialization estimation is done with the motion of the region found in the first step. In the case of rigid objects, motion is estimated for each objects rather than for each couple $(region, object)$, and will be used for each matching couple having the same object. After this motion refinement, splitting of regions is made and areas are affected as presented in the previous section.

In order to limit artifacts in the handling of small detected areas, regions are split only if occluding areas have a significant size, otherwise region is entirely affected to the object it mostly projects on (useful for the case 4 presented in section 3.2 and for small occlusion areas).

The development of this two-step classification permits to deal with regions projecting on several objects. Finally this property allow to use coarser spatial segmentation and to limit artifacts due to the bias in the motion estimation.

4. RESULTS

The proposed modified backward tracking algorithm has been tested on several video sequences with different type of objects, rigid and non-rigid. For the spatial segmentation, we used the segmentation algorithm presented in⁹. Motion compensation is performed using affine motion model estimated with a multi-resolution scheme and relaxations.

Figure 7 shows steps of the backward classification. The spatial segmentation in (b) is backward projected on the object masks defined in (a). Division of regions in affectation areas is presented in (c), and finally in (d), the final tracked segmentation. In this last images, uncovered areas detected have been left. These areas are then well affected using the classification based on the object their spatial regions mostly projects on.

If we look closely to spatial segmentation (b), we can observe that if no splitting was performed, artifacts in the segmentation would appear. For example, if we look at the spatial segmentation, under the wrist, part of the hand is attached with a part of the background, and part of the other hand (at the bottom right of the image) is also missing. Thanks to the further splitting of regions, these artifacts are corrected and are no more visible in the final mask. On this example, we can observe the accuray in the boundary location, and the ability of the proposed algorithm to deal with large motion thanks to the use of multi-resolution motion estimation (see the motion of the ball).

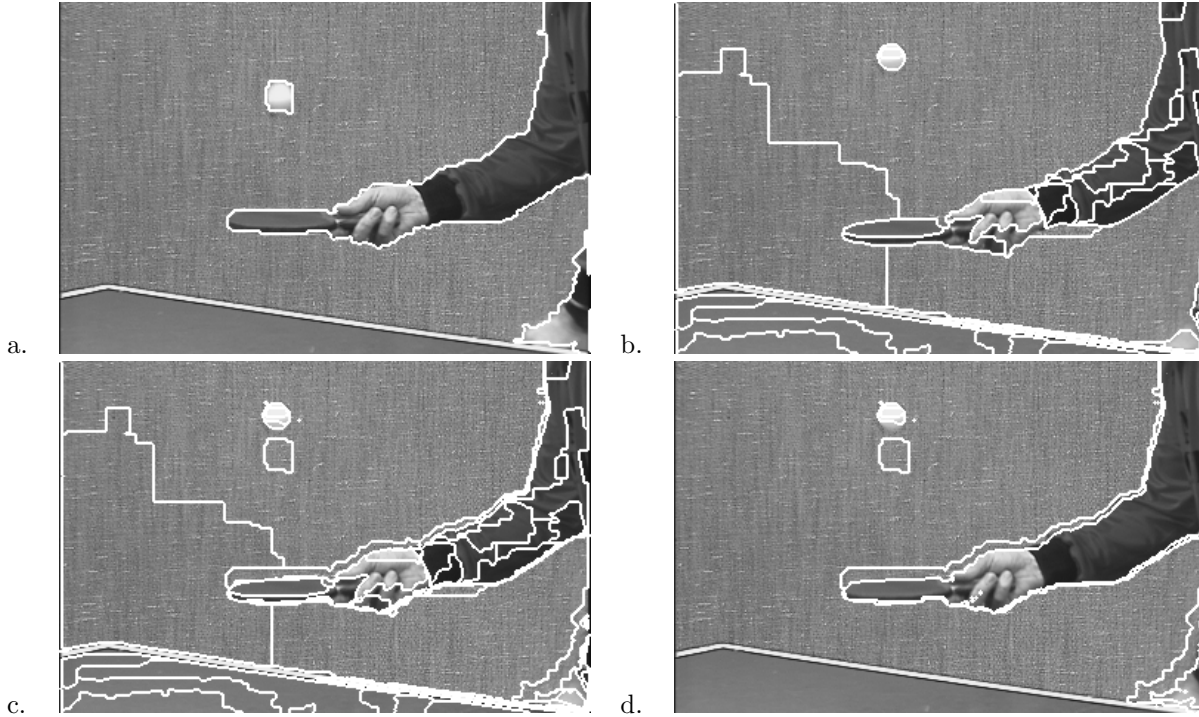


Figure 7. Division of the spatial regions in the backward projection for sequence Table-tennis. Object mask on frame 1 (a), spatial segmentation of frame 3 (b) and further divisions after backward projection (c), final labeling (d).

Figure 8 show results obtained on the tracking of an object on the Foreman sequence. The initial mask was set manually on frame 61 in order to define the face of the man [§]. So two objects were defined, face of the man and a complex object composed of the background and parts of the man. Tracking results show good localization of the boundaries although motion is quite important on this part of the sequence and that the background object is not consistent (composed of the scene background, helmet and shoulders of the man). In this example, each frame was backward projected on frame 61. Artifacts are appearing for frames that are far away from the reference mask since some spatial regions do not have any correspondent in the reference frame (see artifacts in the upper part of frame 99, figure 8).

5. CONCLUSION

In this paper we have presented an extension of the backward tracking technique for video objects. This extension aims at reducing the artifacts appearing in previous proposed approach. This is done by taking explicitly into account occluding areas and the possible errors in spatial segmentation and motion estimation used. The final proposed tracking algorithm show good results of tracking even in the presence of complex motion.

In the future, extension of this work will be done in order to take into account new appearing objects by the use of a detection technique. This extension could also lead to the self definition of the object mask for a sequence. A complete object segmentation algorithm would then be conceivable. Treatment of split areas in a region could also be improve by using global classification of these areas or using markov random fields to define more properly their boundaries. Treatment of uncovering areas could also be improved by using information of next frames.

[§]Previous frames have been left since few motion is present in the beginning of this sequence

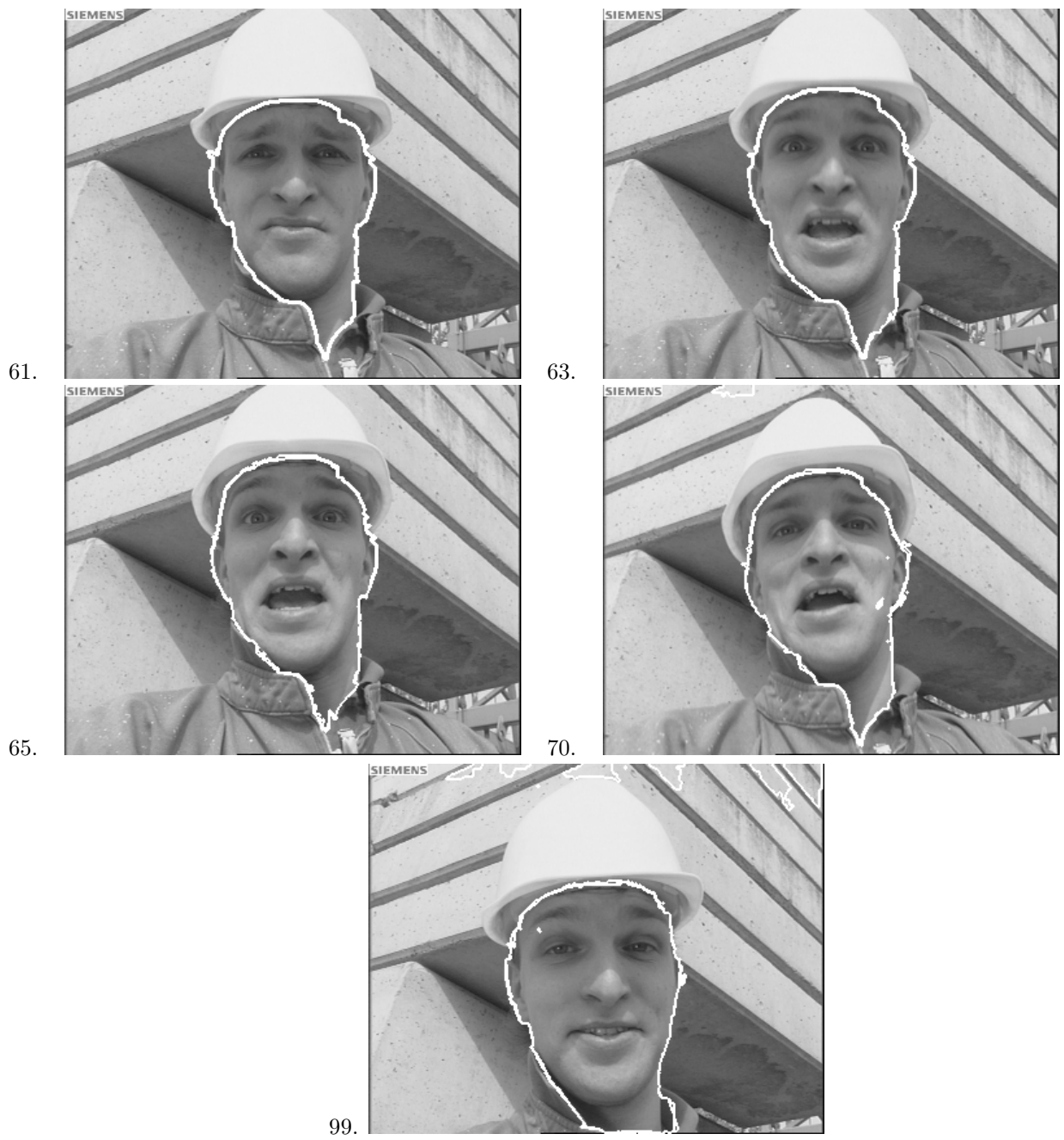


Figure 8. Results obtained for the tracking of the foreman face (frames 61, 63, 65, 70, and 99).

REFERENCES

1. R. Qian and I. Sezan, "Video background replacement without a blue screen," in *ICIP'99*, (Kobe, Japan), Oct. 1999.
2. P. Salembier, L. Torres, F. Meyer, and C. Gu, "Region-based video coding using mathematical morphology," in *Proceedings of the IEEE*, IEEE, ed., vol. 83, pp. 843–857, IEEE, June 1995. Special Issue on Digital Television.
3. L. Bonnaud, S. Pateux, and C. Labit, "Multiple objects tracking for efficient motion-based segmentation coding using a temporal prediction," in *Proc. of PCS 97*, pp. 125–128, (Berlin, RFA), Sept. 1997.
4. R. Castagno, T. Ebrahimi, and M. Kunt, "Video segmentation based on multiple features for interactive multimedia applications," *IEEE Transactions on Circuits and Systems for Video Technology, Special issue on "Image and video processing for emerging interactive multimedia services"* **8**, pp. 562–571, Sept. 1998.
5. F. Marques and J. Llach, "Tracking of generic objects for video object generation," in *ICIP'98*, pp. 628–632, (Chicago, USA), 1998.
6. C. Gu and M. Lee, "Semantic video object tracking using region-based classification," in *ICIP'98*, pp. 643–647, (Chicago, USA), 1998.
7. D. Gatica-Perez, M. Sun, and C. Gu, "Semantic video object extraction based on backward tracking of multivaluated watershed," in *ICIP'99*, (Kobe, Japan), Oct. 1999.
8. F. Marqués and C. Molina, "Object tracking for content-based functionalities," in *VCIP'97*, SPIE, ed., vol. 3024, pp. 190–199, (San Jose, California), Feb. 1997.
9. S. Pateux, "Spatial segmentation of color images according to the MDL formalism," in *ISIVC'00*, (Rabat, Morocco), Apr. 2000.