ADAPTIVE MORPHOLOGICAL TIME STAMP SEGMENTATION BASED ON EFFICIENT GLOBAL MOTION ESTIMATION

Julián Fiérrez, Luis Salgado
{jfa,lsa}@gti.ssr.upm.es
Grupo de Tratamiento de Imágenes
E.T.S. Ingenieros de Telecomunicación

Enrique Navarro
enavarro@inef.upm.es
Laboratorio de Biomecánica
Instituto Nacional de Educación Física

Universidad Politécnica de Madrid – E-28040 Madrid – Spain

ABSTRACT

An efficient segmentation technique of regular shaped sudden temporal intensity changes in image sequences is presented. It is based on morphological analysis of binarized difference images, and its parameters are dynamically modified according to an efficient estimation of the global motion of the scene. Motion analysis is carried out with the combination of a multiresolution flow estimation technique and a closed formula over the resulting motion vectors based on a simplified motion model. Specially suited for applications demanding a temporal reference, the new segmentation technique will be described in the context of a time stamp detection application.

1. INTRODUCTION

The determination of temporal references in sequences of images for synchronization, editing or storage purposes becomes a fundamental problem in fields as audiovisual postproduction and computer vision. The video community tends to use time codes registering them in the same storage medium in which the video signal will be recorded. In computer vision, it is common to acquire a scene from different points of view as an entry point to analyze it. In that case, it is necessary to know the temporal relationship between frames of sequences captured by different cameras, information that is known during acquisition time but is usually lost if storage is carried out prior to processing. To solve automatically that temporal relationship using only the visual information present in the different sequences, a time stamp in each of them is required. There is no common agreement on how to generate and detect that time stamp, and custom solutions are usually employed. The time stamp solution can also solve simple editing problems more economically than complex time codes, making the costs of specialized hardware unnecessary. So, it is a valuable technique with no general solutions, not even design principles.

Particularly in our application, time stamps appear as abrupt locally placed light transitions on the left and right corners of the images, generated by a circuit external to the camera. Figure 1 shows two consecutive frames of a gymnastics sequence, one prior to the time stamp appearance (upper) and the other one holding the light circle time stamp close to the upper right corner.

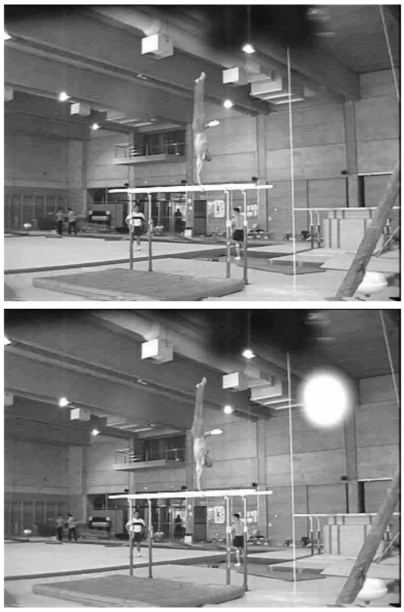

Figure 1. Two frames between which a time stamp is activated: not yet activated (upper), already activated (lower).

Here a new segmentation technique is presented to detect not only the above-mentioned time stamps but also any kind of temporal reference for synchronization or editing if it consists on a regular shaped sudden intensity change through the sequence. Our approach is based on morphological analysis [1] of binarized difference images. To take into account variations of the shape, size and position of the time stamps due to camera movements, processing parameters are dynamically modified according to an efficient estimation of the global motion of the scene. Motion analysis is carried out with the combination of a multiresolution flow estimation technique [2], and a closed formula computed on the resulting motion vectors [3].

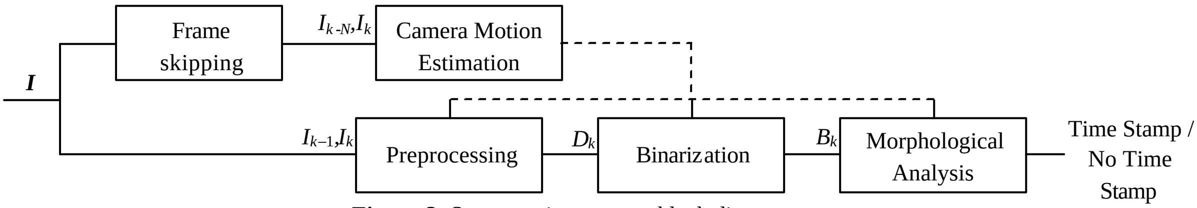

Figure 2. Segmentation strategy block diagram.

2. SEGMENTATION STRATEGY OVERVIEW

A block diagram of the proposed segmentation approach is presented in Figure 2. The input of the system is a sequence $\boldsymbol{I}$ of images I_k where k represents the temporal index. The result of the algorithm is a flag indicating the activation or deactivation of a time stamp between two consecutive frames I_{k-1} and I_k , together with, in case of positive detection, the time stamp location and shape properties.

2.1. Preprocessing

This phase takes two consecutive frames I_{k-1} and I_k and produces a grayscale difference image D_k where the time stamp can easily be detected.

The prior assumption of our segmentation technique is that time stamps to be detected must imply a sudden temporal change of the intensity values in a regular shaped region of the sequence. So, we first take the magnitude of the difference between input images.

To highly improve computational efficiency, processing is restricted to those areas of the image where the time stamp is expected to be located (*areas of interest).* In our application, time stamps are located close to the upper corners of the image. Therefore, the selection procedure indicated in Figure 3 has been used. In it, a reduced image (right) is generated joining the highlighted left and right upper areas of the original image (left). Further processing is carried out on this reduced image.

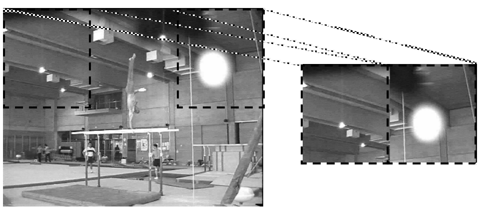

Figure 3. Area of interest selection.

2.2. Binarization

In order to explain our binarization scheme, let us consider a fixed percentage P of the total number of pixels of the input image D_k. We postulate that, with the proper selection of P, the set of higher intensity image pixels covering the P percentage of the image will have: shape properties similar to that of the original time stamp in the case it has appeared or disappeared between the two input images, and a set of irregular distributed

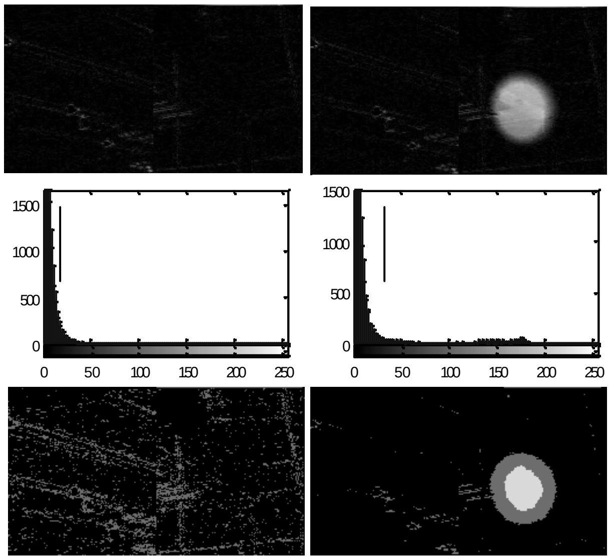

Figure 4. (From top to bottom) Difference image, its histogram and erosion (light gray) over thresholded image (dark gray). Right column shows the activation of a time stamp and the left one shows no stamp.

pixels in other case. Our aim is to calculate that P value (in a set up stage) and binarize the input image D_k applying a histogram-based thresholding technique taking it into account.

In the set up stage, the output binary image B_k is obtained applying a global threshold T_0, which is the minimum value restricted to: $T_0 \geq (m + s/2)$ being m and s respectively the mean and standard deviation of the histogram of the input image D_k. This empirical selection of the threshold (taking into account the following stages) has reported good results in practice detecting usual time stamps. The set up stage will finish in the first time stamp detection. At this point, P will be calculated as the ratio between the number of pixels above the T_0 threshold and the total.

Once P is calculated, the thresholding is carried out working with the cumulative histogram $H_c(i)$ of the input image D_k, that was previously scaled to a unity maximum value. The global threshold T for the image D_k is selected as the minimum value restricted to $1-P \leq H_c(T)$.

2.3. Morphological Analysis

This phase takes as input the binary image B_k. It is eroded and the transition between I_{k-1} and I_k is considered to hold a time stamp if the resulting image is non-zero. The segmented time stamp will be the dilation of the resulting image by the same structuring element used for the erosion.

The suggested structuring element's shape is the same of that of the time stamps to be detected. The size must be the largest possible in order to correctly detect false time stamps but smaller enough to detect all possible ones. It is selected in an off-line configuration process where the minimum size time stamp is processed and eroded by a set of increasing size structuring elements until the resulting image is zero. In our application, the minimum time stamp size corresponds to the case of acquiring under minimum zoom conditions.

As an example of the whole segmentation process, images originated in our application are depicted in Figure 4. The upper left and right are two preprocessed images D_1 and D_2 where no time stamp activation/deactivation has occurred and time stamp activation/deactivation has occurred respectively. Below each one we represent its histogram and the resulting threshold (as an arrow) applying the setup technique ($m+s/2$). Finally and below each histogram the resulting thresholded images B_1 and B_2 are indicated in dark gray in which the erosion result is overprinted in light gray. As can be observed, the resulting erosion is non-zero only in the right one, where time stamp is detected.

3. ADAPTIVE STRATEGY

Under the presence of camera motion or zoom, and if temporal references are generated by an external circuit, time stamps are likely to vary their properties in time. This is the case of our application, where zoom changes time stamps size, an effect that is compensated with a motion based adaptive strategy.

3.1. Motion Model

The model to fit for motion estimation is that of a fixed camera that can rotate about its three axis with angles a (pan), b (tilt), and g, and can vary its focal length from f to $s \cdot f$ as used by Tan et al. in [4]. In case of motion, they demonstrate that the projected locations of a three dimensional point between two images I_{k-N} and I_k are related by the six parameter projective transformation:

$$x(k) = \frac{p_1 x(k-N) + p_2 y(k-N) + p_3}{p_5 x(k-N) + p_6 y(k-N) + 1}$$
$$y(k) = \frac{-p_2 x(k-N) + p_1 y(k-N) + p_4}{p_5 x(k-N) + p_6 y(k-N) + 1} \qquad (1)$$

where $w(k-N)=[x(k-N),y(k-N)]^T$ and $w(k)=[x(k),y(k)]^T$ are the projected locations of the three dimensional point on image I_{k-N} and I_k, and p_1, p_3 and p_4 can respectively be approximated by s, $-s \cdot a \cdot f$ and $s \cdot b \cdot f$, assuming a small rotation change. Let us consider the following simplifications: $p_5 = p_6 = 0$ (i.e. perspective distortion effects are minimal), and $p_2 = 0$ (i.e. the camera does not rotate about the axis of the camera lens). It can be shown [3] that p_1, p_3 and p_4 can then be estimated, providing a set of noisy motion vectors $w_i(k-N) \rightarrow w_i(k)$, by the closed formulas:

$$p_1^* = \frac{\sum_i (w_i(k) - \overline{w}(k))^T (w_i(k-N) - \overline{w}(k-N))}{\sum_i \left| w_i(k-N) - \overline{w}(k-N) \right|^2}$$
$$\begin{bmatrix} p_3^* / p_1^* \\ p_4^* / p_1^* \end{bmatrix} = \frac{\overline{w}(k)}{p_1^*} - \overline{w}(k-N) \qquad (2)$$

where:
$$\overline{w}(k) = \frac{1}{M} \sum_i w_i(k)$$

M is the total number of motion vectors and p_1^*, p_3^* and p_4^* are the least mean squares estimated values.

3.2. Motion Analysis

The motion analysis is carried out by a non-dense multiresolution flow estimation technique.

A set of feature points easy to track will be first selected based on Shi-Tomasi's criterion [2]. This set should be uniformly distributed over the entire image, so a minimum distance criterion between them is also applied. These points are tracked through the sequence generating the motion vectors $w_i(n-1) \rightarrow w_i(n)$. If the tracking goes wrong for a selected feature point, then it will be discarded. If a fixed percentage of initial points are lost this way, then new features will be calculated to complete another uniformly distributed set.

The tracking is carried out using the Lucas-Kanade iterative algorithm [2]. Our implementation work on a multiresolution pyramid that allows for even large displacements between images (and so allows skipping frames for motion analysis).

3.3. Parameter update criteria

In our segmentation scheme, the adaptive strategy is carried out modifying three parameters: the area of interest (the portion of the input images to be processed), the P value (percentage of pixels to be thresholded) and the size of the structuring element. The estimated values p_1^*, p_3^* and p_4^* in (2) (i.e., s, $-a \cdot f$ and $b \cdot f$) help to time adapt the above-mentioned parameters.

Here we will demonstrate how the zoom effect was compensated in our application. Considering $Z(k)$ as the zoom factor between frames $k-1$ and k the following criteria was applied: A) The sizes of the area of interest and the structuring element were scaled using a factor equal to $Z(k)$. B) Being $r(k)$ and $Area(k)$ respectively the radius and the area of the time stamp at time k and according to the fact:

$$Area(k) \approx p\,(r(k))^2 \approx p\,(Z(k) \cdot r(k-1))^2$$
$$\approx Z(k)^2 \cdot Area(k-1)$$

the update criterion used for the percentage P of pixels to be thresholded was:

Figure 5. Three frames of a zooming sequence marked with the shutting LED circuit together with segmented time stamps (black grids) and features used for motion estimation (white circled black points).

$$P(k) = Z(k)^2 P(k-1)$$

where $Z(k)$ is calculated from p_1^* using the following formula (that takes into account the different rate between motion estimation and time stamp segmentation due to frame skipping):

$$Z(k) = (p_1^*)^{1/N}$$

4. RESULTS

Tests have been carried out using real grayscale CIF sequences at 50 frames/sec. (combining zooming, panning, moving objects and illuminations changes). Time stamps in almost all of them were correctly segmented. Figure 5 shows an example where three time stamped images of an incremental zooming sequence (from 1x to 3x) are represented. The detected time stamps (as black grids in the Figure) and the tracked features (as white circled black points in the Figure) are also overprinted. As can be observed, time stamps are correctly detected and segmented thanks to the adaptive strategy implemented.

In order to quantify the error introduced by the simplified motion model, we have estimated the zoom value not only using (2) but also by least mean squares fitting the general transformation (1). Results for the zooming sequence presented in Figure 5 are shown in Figure 6. In the left part it is shown the zoom estimated using (1). The results using (2) are so similar that a separate graph (right) with its relative difference is included for comparison. It can be seen that, for this sequence, the relative difference is never higher than ±0.15%.

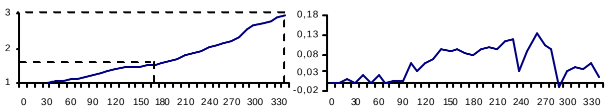

Figure 6. Estimated zoom using the non-simplified motion model (1) considering initial zoom as 1x (left) and its relative difference (%) to that using the simplified motion model (right), both indexed by the frame number.

We have also tested synthetic noisy images. An example is shown in Figure 7. In the upper part it can be seen an area of interest to which a gaussian noise with s = 25 has been added (left). In the right part, the resulting thresholded difference image together with its erosion is represented. The time stamp is correctly detected as the algorithm is specially robust to that kind of noise. In the lower part of Figure 7, results are shown for uniformly distributed impulsive noise. To improve the segmentation robustness in that case, a 3x3 median filter has been applied at the end of the preprocessing stage.

Real time operation on a conventional PC was achieved processing only in the area of interest, skipping 9 of 10 frames for motion estimation (so N = 10) and considering a square shaped structuring element.

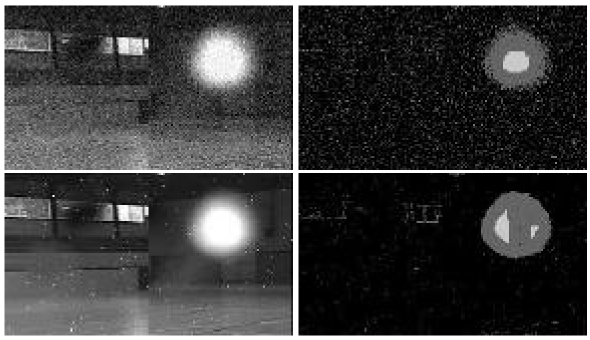

Figure 7. Test of gaussian (upper) and impulsive noise (lower). Noisy areas of interest are represented (left) and also their corresponding erosion over thresholding (right).

5. REFERENCES

[1] R. M. Haralick, S. R. Sternberg, and X. Zhuang, "Image analysis using mathematical morphology", *IEEE Trans. PAMI*, Vol. 9, nº 4, pp. 532–550, 1987.

[2] J. Shi and C. Tomasi, "Good features to track," *IEEE Conference on CVPR*, pp. 593–600, 1994.

[3] Y. P. Tan, D. D. Saur, S.R. Kulkami, P.J. Ramadge. "Rapid Estimation of Camera Motion from Compressed Video with Application to Video Annotation". *IEEE Transactions on CSVT*, Vol. 10, nº 1, pp. 133–146, Feb. 2000.

[4] Y. P. Tan, S. R. Kulkarni and P. J. Ramadge, "A New Method for Camera Motion Parameter Estimation", *IEEE International Conference on Image Processing*, Vol. 1, pp. 406–409, Oct. 1995.