ADAPTIVE SEGMENTATION FOR GYMNASTIC EXERCISES BASED ON CHANGE DETECTION OVER MULTIRESOLUTION COMBINED DIFFERENCES

José M. Cobo, Luis Salgado, and Julián Cabrera

{jmc,lsa,jcq}@gti.ssr.upm.es
Grupo de Tratamiento de Imágenes
E.T.S. Ingenieros de Telecomunicación - Universidad Politécnica de Madrid - Spain

ABSTRACT

A new adaptive segmentation strategy is proposed to accurately segment gymnasts in sport sequences. It is based on a Markov Random Fields (MRF) change detection analysis operating on a multiresolution combination of static and dynamic image differences. After a morphological analysis of the segmented masks, it incorporates estimated motion information in the area of interest to improve the efficiency of the segmentation process. Although presented in the particular context of gymnastic exercises, the new segmentation strategy could be applied to other applications where moving objects on a quasi-static background need to be segmented.

1. INTRODUCTION

Biomechanics is a science that deals with human body analysis with different goals, such as sport or medical ones; it becomes of special interest whenever a precise modeling of a complex mobile system is needed (like scientific simulation or artificial animation). The introduction of illumination tags or markers on the gymnasts' bodies simplify the segmentation step as simple tracking algorithms can be directly implemented to analyze their evolution. Nevertheless, when those intrusive systems can not be used, a supervised analysis needs to be carried out. Points of interest are manually marked on the recorded sequences of the gymnastic exercises for further motion analysis. To simplify and improve the accuracy of this process, a segmentation of the gymnast along the sequence becomes fundamental.

Figure 2 (a) presents a sample of the type of gray level sequences under consideration. Several factors contribute to make particularly difficult the segmentation process: the lack of color information (which would ease significantly, as in other sports analysis approaches, the segmentation process), the variability in size of the object of interest with respect to the image size, and the wide range of movements of the gymnast in its exercise. For all these reasons, a high adaptability from the segmentation strategy to be designed is demanded.

Here a new adaptive segmentation strategy is proposed to accurately separate the gymnast in the scene to ease further object tracking and motion evolution analysis. It is based on a MRF change detection operating on a multiresolution combination of static and dynamic image differences, and a morphological analysis of the segmented masks. It incorporates estimated motion information in the area of interest to improve the efficiency of the segmentation process. Flexibility is achieved through the dynamic adaptation of segmentation parameters to improve the accuracy of the results to deal with variable object sizes and motion evolution.

2. SYSTEM OVERVIEW

A block diagram of the proposed segmentation approach is presented in Figure 1. The input of the system is a sequence ***I*** of images, being *I(n)* the image corresponding to the temporal instant *n*. The result of the algorithm is a set *{R}* of masks showing the segmented object along the sequence.

2.1. Pyramid generation

The variability in the illumination conditions and the noise present in the acquired images produce an over-detection of changing areas when segmentation is computed directly on sequences at full resolution. To overcome this limitation, a multiresolution approach has been implemented. A hierarchy of subsampled images (the pyramid) is generated following a simple quad-tree approach [1], where each tile of 2x2 pixels at one level of the pyramid is averaged to generate one pixel at the lower resolution image. The input to the module are full resolution images $I_0(n)$, also identified in the present text as *I(n)*. A set of subsampled images $\{I_1(n), I_2(n), \ldots, I_k(n), \ldots, I_L(n)\}$ is then generated, being $I_L(n)$ the lowest resolution images of the pyramid. Thus, the subsequent segmentation steps work on the lowest resolution sequence $I_L(n)$. Figure 2 (b) shows the change detection result operating at the first level of the pyramid. As it can be observed, the main detected object corresponds to the gymnast, and other small noisy detections appear in the background.

2.2. Motion estimation

Motion analysis is carried out by a non-dense multiresolution flow estimation technique. A set of features points easy to track will be first selected based on Shi-Tomasi's criterion [2]. These points are tracked generating the motion vectors $\omega_t(n-1) \rightarrow \omega_t(n)$. Tracking is carried out using the Lucas-Kanade iterative algorithm [3], working on a multiresolution pyramid. It operates on three subsequent images *I(n-2)*, *I(n-1)* and *I(n)*.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

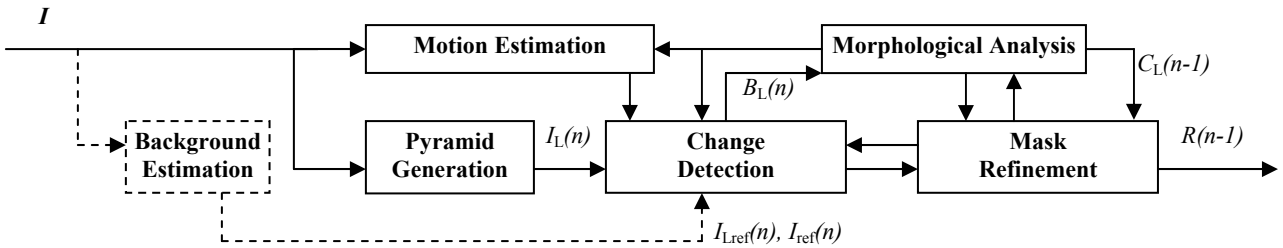

Figure 1. Block diagram of the proposed segmentation approach

Motion estimation is computed for every image on the area covered by the projection to full resolution of the rectangle bounding the detected object mask $M_L(n-1)$, enlarged by a security margin which is dynamically modified according to the size of the object. The result is a set of motion vectors $\omega_i(n-2) \rightarrow \omega_i(n-1)$ which shows the direction and size of the displacement. This motion information is computed locally to the area where the object of interest has been detected in previous images. It is used by the change detection stage as a prediction to dynamically reshape the region where the object of interest is to be detected.

2.3. Change detection

To successfully segment the moving objects, assuming a quasi-static background, a change detection strategy based on the computation of image differences is proposed. It is carried out through the computation of a combined static/dynamic difference image that is used to detect the changed pixels through an adaptive MRF based change detector.

2.3.1 Background estimation

To compute the static difference image, a static background reference image is needed. When this reference is not available from the beginning of the analysis, it must be built using a background registration technique similar to that described in [4].

To overcome illumination variability issues, the new pixels belonging to the background are computed by means of a recursive temporal filtering defined by the following expression ($\beta \in [0.5,0.95]$), applied only to pixels already known to belong to the background:

$$I_{Lref}(n) = \beta I_L(n) + (1 - \beta) I_{Lref}(n-1) \quad (1)$$

2.3.2 Combined difference image computation

A combined difference image $D_L(n)$ is generated from the combination of static and dynamic difference images [5]:

$$D_L(n) = D_{Lstat}(n) + D_{Ldyn}(n) \quad (2)$$

where $D_{Lstat}(n)$ corresponds to the static difference image obtained subtracting the current image $I_L(n)$ from the estimated reference image, $I_{Lref}(n)$, and $D_{Ldyn}(n)$ corresponds to the dynamic difference image obtained subtracting the current image $I_L(n)$ from the previous one $I_L(n-1)$. To avoid erroneous static detections when the background is not fully reconstructed, undetermined background pixels are substituted by their corresponding $I_L(n)$ pixels. Therefore, for those pixels the static difference image term is cancelled.

2.3.3 Changing areas determination

A new adaptive implementation of a pixel oriented change detector based on the nine thresholds algorithm described by Aach and Kaup [6] is proposed. Pixels in the binary image $B_L(n)$ are labeled as static (not changed) or dynamic (changed) according to the relationship between the pixels in a square neighborhood defined by N_w (size of the neighborhood to be analyzed for each pixel), a set of nine pre-determined thresholds, the error probability α (to label a static pixel as dynamic and viceversa), and the so-called "potential value" B [6]. The nine thresholds, one corresponding to the overall image information, and eight obtained through a bidimensional Markov Random Field modeling of the pixel neighborhood [7], are applied over the difference image $D_L(n)$.

The decision rule on a pixel of image $D_L(n)$ (pixel i changed or unchanged) of this algorithm is:

$$\overline{\Delta_i^2} > t(j) \Rightarrow \text{Pix. } i \text{ chang.} \quad \overline{\Delta_i^2} < t(j) \Rightarrow \text{Pix. } i \text{ unchang.} \quad (3)$$

with the left-side term being the normalized square sum of the pixels in a window w_i of size N_w centered at the pixel i of $D_L(n)$:

$$\overline{\Delta_i^2} = \frac{1}{\sigma_0^2} \sum_{k \in w_i} D_L^2(n,i) \quad (4)$$

The right-side term of (3), t(j), is a vector containing the referred nine thresholds. The so-called "anchor threshold", $t_S = t(4)$, is determined by the error probability parameter α, and the other eight threshold values are computed with the expression:

$$t(j) = t_S + 16*B - 4*B*j \quad (5)$$

where $j \in [0,8]$ is the number of changed pixels present in a 3x3 window centered at pixel i of the generated change mask. When j pixels in that window have been marked as changed by the previous iteration of the algorithm, the threshold t(j) is selected, and the described decision rule can be applied.

The normalization factor σ^2_0, presented as a constant noise measure in the referred works, is dynamically modified in our implementation to take into account the evolution of the results of the segmentation process for previous images in the analyzed sequence. This normalization factor is adapted according to the size of the mask obtained from the previous image: a very small segmented object imply lower values of $D^2_L(n)$ in the window w_i, as long as background pixels (lower change value) could be more probable than mask pixels inside w_i due to the small size of the object. This may cause the numerator in (4) to decrease,

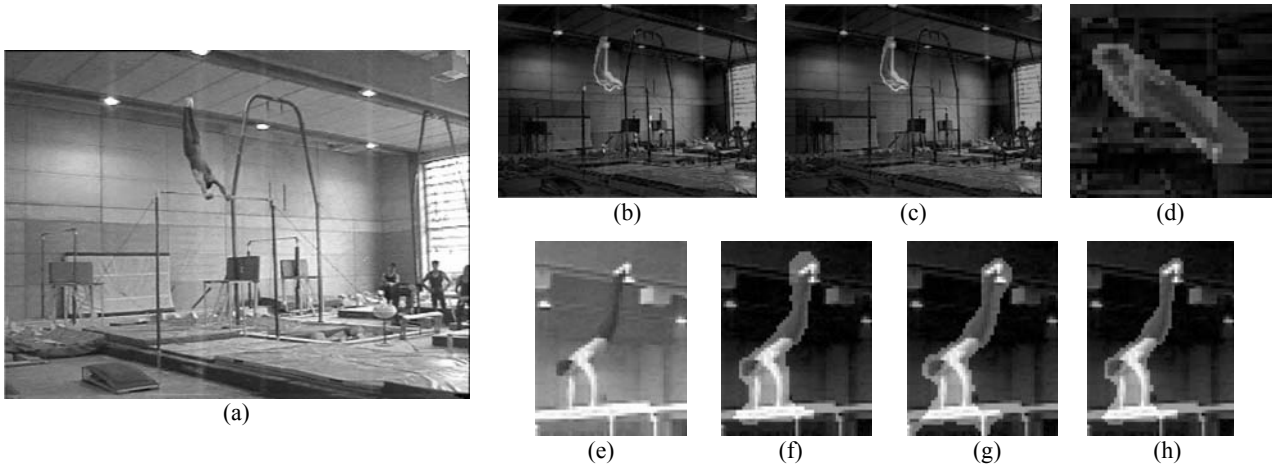

Figure 2. Change detection and morphological analysis (details).

keeping the denominator constant in a way that would cause wrong labeling of changed pixels as unchanged as they would not reach the threshold. Therefore, parts of the object would be missed. Here the denominator is dynamically modified to avoid that.

Additionally in our proposal, the innovation using both dynamic and static differences to generate the image to be analyzed improves significantly the performance results of the algorithm: it allows achieving an extra reliability in the change areas determination process while it reduces the possibility of missing the object if it stops or moves too slowly, as long as less homogeneous regions (with "holes" in zones with a homogeneous gray level) could be produced if only dynamic differences were used [5].

Assuming a quasi-static background, the output of the whole change detection stage is the binary image $B_L(n)$, which is formed by a set of regions corresponding to the moving areas in the analyzed image.

2.4. Morphological analysis

To generate compact segmented areas as those likely to correspond to relevant objects, a morphological analysis is required.

A closing operation is carried out to achieve more compactness in the change areas, as well as to connect occasionally disconnected ones belonging to the same moving object. Due to the expected smooth contours of the object of interest, a circular structuring element with constant size is used.

The result of the closing operation is a set of homogeneous regions corresponding to moving objects in the region of interest. Particularly, for the sequences under consideration, as only one object of interest (the gymnast) is appearing in the scene, and the acquisition is focused on its evolution, the largest detected region is selected. Figure 2 (c) shows the result of the morphological analysis on the change detection presented in (b), with small regions already removed.

The output is a mask image $M_L(n)$ corresponding to the segmented object of interest. $M_L(n)$ is also used by change detection and motion estimation stages to dynamically adapt their processing areas based on the size of the present mask. Figure 2 (d) shows the detected mask for a sample image of the vaulting horse exercise, working with L=2 because of the much larger size of the object. As it can be seen, there is apparently a loss of accuracy in the algorithm, due to be working on a level L>1. Accuracy is however recovered later, after the mask refinement stage, as it can be seen in Figure 3 (e)-(g).

As long as the mask image $M_L(n)$ contains motion information of two images $I_L(n)$ and $I_L(n-1)$, further processing is needed. Let us have three images of the sequence $I(n-2)$, $I(n-1)$ and $I_L(n)$. $M_L(n-1)$ and $M_L(n)$ are the object masks obtained from $I_L(n-2)$, $I_L(n-1)$, and $I_L(n-1)$, $I_L(n)$ respectively. The common part of both masks corresponds to the desired object mask $C_L(n-1)= M_L(n) \cap M_L(n-1)$, as it reflects motion information belonging exclusively to the image $I_L(n-1)$ Figure 2 shows these masks for sample (e) (working with L=1, as long as the object is small related to the image size): (f) shows $M_L(n-1)$, with motion information of (e) and its previous image; (g) shows $M_L(n)$, with motion information of (e) and its following image, and (h) shows the intersection of both, $C_L(n-1)$, corresponding to the object mask of (e).

2.5. Mask refinement

$C_L(n)$ corresponds to the segmented object computed on image $I_L(n)$. To obtain a refined mask image $R(n)$ corresponding to the full resolution image $I(n)$, a projection of this mask to higher resolution levels is carried out [1]. This projection results in a loss of accuracy in the contours of the mask, that is reduced applying a new iteration of the change detection and morphological stages restricted to the set of the pixels belonging to the refined object mask at full resolution.

Let $H(n)$ be the set of boundary pixels of the full resolution mask image $C(n)$, obtained as the logical xor operation of $C(n)$ and its erosion $E(n)$. Pixels belonging to $E(n)$ are directly considered as part of the final object mask $R(n)$. Change detection operates over $H(n)$ and $H(n-1)$. The mask image result of the stage is morphologically opened and closed to smooth the contour, obtaining the boundary mask $T(n)$. Then, the mask containing the change information of the full resolution images $C(n)$ and $C(n-1)$ is obtained joining these masks: $G(n)= E(n) \cup T(n)$. A last morphological and region analysis is needed to identify isolated small regions, that then are rejoined to the

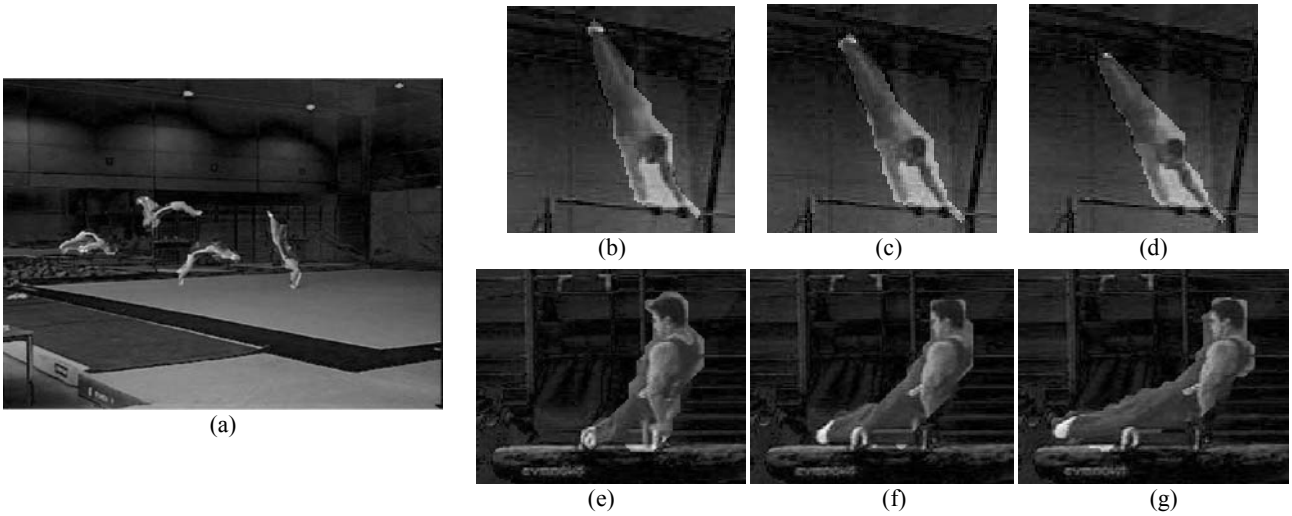

Figure 3. Final segmentation results

mask, identified as S*(n),* by means of a morphological closing.

After this, again the intersection of two masks is required to achieve the final result, as stated before for the *L* level masks: *R(n-1)= S(n)* $\cap$ *S(n-1)*. *R(n)* would be the full resolution refined mask corresponding to the input image *I(n).*

3. RESULTS

To generate the reference image, $I_{Lref}(n)$, β is set experimentally to 0.95, setting $\alpha = 5 \times 10^{-4}$ and B=2.25. The evaluation mask used is 3x3 (N_w=9) instead of other suggested values in the literature, as accuracy in the detected masks is improved while computational requirements are reduced. Adaptability in this phase is achieved dynamically modifying the value of σ^2_0 to be 1/25 of the object mask's size already computed, with a maximum value of 10 when working at L=2 to avoid elimination of relevant parts of the object due to the smaller image resolution. At the mask refinement stage, σ^2_0 for change detection is set to 1/50 of the unrefined mask's size.

Figure 3 shows final segmentation results for different exercises. For the floor sequence, a case in which the gymnast moves all around the image with very fast movements, (a) shows the segmentation results at different time instants. The masks obtained have a very high degree of accuracy for such a small (L=1) moving object, with only some expected minor erroneous detections between arms and legs; this is due to the selected morphological closing strategy which keeps the thinnest parts of the object. The opposite situation is presented in Figure 3 (b)-(d), where slow movement of the small (L=1) gymnast is present. Accuracy is fine enough, though there exist still some small zones detected as changed which should be further refined with strategies that ensure keeping the thin regions (arms) of the object. Finally, segmentation results when the object size is large (vaulting horse exercise, L=2) are presented in the same figure (e)-(g), where the mask refinement stage has significantly improved the mask accuracy.

4. CONCLUSIONS

A new adaptive segmentation strategy based on MRF change detection analysis operating on a multiresolution combination of static and dynamic image differences has been proposed. Although presented in a particular context, the new segmentation strategy could be used in other applications where moving objects on a quasi-static background need to be segmented.

Further work should include the use of motion estimation information and morphological coherency between subsequent images to complement the morphological analysis to improve accuracy in the segmentation.

5. ACKNOWLEDGEMENTS

This work has been partially supported by the Comisión Interministerial de Ciencia y Tecnología of the Spanish Government under project TIC2002-03692 (DYMAS).

6. REFERENCES

[1] L. Salgado, N. García, J. M. Menéndez, and E. Rendón, "Efficient Image Segmentation for Region-Based Motion Estimation and Compensation", *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 10, Nr. 7, October 2000.

[2] C. Tomasi and J. Shi, "Good Features to Track", *Proc. IEEE Conf. Comput. Vision and Pattern Recogn.*, pp. 593-600, 1994.

[3] B.D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision", *Proc. of Image Understanding Workshop*, pp. 121-130, 1981.

[4] S. Chien, S. Ma, and L. Chen, "Efficient Moving Object Algorithm Using Background Registration Technique", *IEEE Trans. CSVT, Vol. 12, No. 7, pp.577-585. July 2002.*

[5] L. Salgado, N. García, J. M. Menéndez, and E. Rendón, "Dynamic Object Segmentation for Outdoor Analysis", *Proc. Visual Comm. and Image Processing*, pp. 593-603, 2002.

[6] T. Aach, and A. Kaup, "Bayes. algor. for adapt. change detect. in image seq. using Markov Random Fields", *Signal Proc.: Image Communications*, vol.7, no. 2, pp. 147-160, 1995.

[7] R. Chellapa and A. Jain, *Markov Random Fields Theory and Applications*, Academic Press, Inc., 1993.