ACCURATE MOTION INTERPOLATION BY USING A REGION BASED MOTION ESTIMATOR

R. Lancini, M. Ripamonti, S. Tubaro and P. Vicari
CEFRIEL, Politecnico di Milano,
Via Fucini 2, Milano, I-20133, ITALY
Tel. +39 2 23954-209, Fax. +39 2 23954-254
e-mail: rosa@cefriel.it

ABSTRACT

In this paper an effective motion compensated image interpolation techniques is presented. The proposed algorithm has been developed for the interpolation of missing frames in image sequence and it is based on two principal elements. The first is a region-based motion estimation and representation technique, which combines, for each known image, an approximate initial motion field estimate and an initial image over-segmentation (obtained using both luminance and chrominance information) to produce a very accurate affine-model regularized motion field. The proposed algorithm uses a robust identification technique for the estimation of the motion parameters associated to each region. The second important element is the use of an interpolation technique, that starting from the available motion information, defines, for each point of the image to be interpolated a suitable reconstruction strategy taking into account for different situation that can appear (for example a moving object can occlude either a stationary background or an another object and so on). The principal aim of the proposed algorithm is the reconstruction of the missing frames in a *reasonable* way, without introducing significant artifacts and assuring the pleasant representation of object displacement.

1 INTRODUCTION

This article deals with the problem of accurate interpolation of missing frames in image sequences. It is well-known from the literature [1], [4] that this type of interpolation can be implemented in an effective way only by taking into account the motion of objects present in the imaged scene and, therefore, by implementing non-linear interpolation techniques. This is mainly due to the fact that, generally, the frame rate in image sequences does not satisfy the hypothesis of the sampling theorem and, therefore, linear interpolation techniques cannot actually be used. This fact becomes even more critical when a time- subsampled version of an image sequence is being considered and the missing frames must be reconstructed. Region-based motion estimation algorithms are particularly suitable for object-oriented video coding schemes and high-quality motion compensated post-processing of video streams. In fact, region-based motion estimation algorithms can lead to estimated motion fields that exhibit a high degree of coherence inside each region that should represent the perspective projection of a physical object that moves in the imaged scene. One crucial aspect of the region-based motion estimation technique is that it actually is a *circular* problem, which consists of two sub-problems that cannot be solved separately. One problem is relative to accurate segmentation of images into regions of coherent-motion (*objects*) while the other one is that of the estimation of motion parameters for each one of the regions. In principle, the mutual inter-dependency between the two sub-problems would require a joint approach to their solution. However, the complexity of joint estimation algorithms currently makes them unsuitable for real-time applications. This article presents a region-based motion estimation scheme in which the two sub-problems of image segmentation and parameter estimation are treated as separately as possible. In order to do so, two rather complementary information sources, extracted from the input image sequence, are combined: the former is a modestly accurate motion field estimate while the latter is an initial image segmentation into *elementary regions*, which is obtained through a purely intra-frame analysis. Two significant results are presented in this work: the former consists of an algorithm for an accurate description of the vector field that describes the motion between two consecutive images (current one and previous one). This algorithm is also able to segment the current image into regions characterized by congruent motion. The second result consists of an interpolation technique that, starting from the available motion information, defines for each point of the image to be interpolated a suitable reconstruction strategy that accounts for the possible situations that could take place (e. g. a moving object which occludes either a still background or an some other object). These two methods are strictly interconnected because a sophisticated interpolation algorithm can only be effectively implemented when reliable and physical motion information is available. The performance of the entire algorithm can be increased taking into account the time redundancy of the considered sequence. In particular the number and the shape of the regions (objects) should vary slowly along time. Therefore for each new image the segmentation and displacement information - relative to the previously analyzed frame - can be used as a starting point for the analysis. As far as the interpolation algorithm is concerned, the main goal of the interpolation technique is to reconstruct missing frames of an image sequence without introducing significant artifacts while guaranteeing a high-quality representation of the objects that are moving within the imaged scene.

2 THE PROPOSED ALGORITHM OF MOTION ESTIMATION

In the algorithm we propose, the initial image segmentation and motion estimation operations are applied by two parallel functional units on the same video data, as shown in fig. 1.

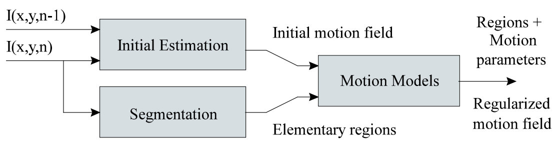

Figure 1: Architecture of the proposed algorithms.

2.1 Initial Estimation

A very efficient, block-based algorithm [6] is used for the initial estimation of the motion vectors between the current image *I(x,y,n)* and the past image *I(x,y,n-1)*; spatial and temporal coherency are exploited both to reduce the computational cost and to increase the coherency of the estimated motion field.

2.2 Segmentation

This stage produces an intra-frame spatial partition of the current image $I(x, y, n)$ into an initial large set of regions, using three steps: image simplification (by anisotropy diffusion), contour extraction (by the morphological gradient), region labeling (by the watershed transformation). Both luminance and chrominance edge information is employed, to avoid segmentation ambiguities.

2.3 Motion Models

The purpose of this block is to identify a set of objects characterized by coherent motion, by suitably grouping elementary regions (coming from the segmentation stage) by an iterative analysis of the available motion information. Each group of neighbouring elementary regions characterised by *coherent* motion vectors is grouped into a unique macro-region and an affine motion model is estimated in order to assign a displacement vector to each region point. The clustering procedure is done, in an iterative way, by performing a robust linear regression (*M-estimation*) [2], [3] on the block-estimated motion vectors. As a matter of fact, robust linear regression represent a reasonable solution for the parameter identification problem when observed data is affected by errors that cannot be modeled as a *white* random noise with *Gaussian* distribution. Since the regression is done on arbitrarily shaped regions, blocking artifacts are eliminated, and a more accurate estimation of the true object motion is achieved even if the initial motion information is noisy. An interesting *side-effect* of the robust clustering and estimation of the motion parameters is represented by the fact that, normally, the points belonging to covered/uncovered areas (*critical zones*) are clearly detected. This happens because the regression of the affine parameters takes into account both the *initial* motion vectors and their corresponding motion compensated luminance differences.

2.4 Region tracking

In order to increase the quality of the motion and the segmentation information, the region shapes and the affine models, obtained from the previous frame, can be used as a starting point for the current frame analysis. At first the intra frame segmentation is carried out as indicated in section 2.2. The merging of the "micro-region" into a region that should represent real object, uses the motion knowledge to update the previous available information. Both motion information and statistical tests on the region shapes - present in the previous and current images - are used to validate the current match between regions into the two images (current one and previous one). When a good match is not found, the system assumes that a scene change is occurred and all the processes are restarted from scratch [7], [8].

3 THE PROPOSED INTERPOLATION TECHNIQUE

A number of algorithms for temporal interpolation is available in the literature, but most of them either rely on the assumption that the background is fixed and the objects never overlap with each other, or use very simple heuristic rules for reducing interpolation artifacts [5]. In the next subsection we will explain how the availability of a region-based motion estimator can lead to a better solution to the temporal interpolation problem. We assume to work with time subsampled version of image sequences version (M indicates the subsampling factor) and to reconstruct the missing images. The step of the interpolation process are the following.

3.1 Region map for past image and intermediate image

We define as a region map m_n a geographical map in which all points belongings to one region are marked with a single value; m_n is obtained by segmentation of the image I_n. If the motion information, associated to each region, is known, then it is also possible to reconstruct this map also for the image I_{n-M} (m_{n-M}), where M is the time-subsampling factor, through motion compensation, that produces the multi-assigned pixel in m_{n-M} belonging to the uncovered regions. If two or more pixels in m_n are assigned to the same pixel in m_{n-M}, we choose the one that minimizes the motion-compensated luminance difference between images I_n and I_{n-M}. Furthermore, if we assume that shape transformations are continuous over time (which corresponds to the uniform linear pixel motion assumption), then it is possible to define a depth-priority criterion for overlapping regions, which permits to obtain the map m_{n-k} for $0 < k < M$, relating to the intermediate image I_{n-k}. This map contains the covered/uncovered zones and the moving objects in the intermediate image.

3.2 Motion models in intermediate image

As a consequence of motion continuity, it can be shown that, if an affine transformation describes the shape changes of a region between I_n and I_{n-M}, then there exists an affine transformation that exactly describes the shape changes of the same region between $I_{n-k} - I_n$ and between $I_{n-k} - I_{n-M}$.

3.3 Reconstruction of pixel data in intermediate image

Once the region map m_{n-k} and the spatio-temporally transformed motion models for all regions are known, the reconstruction of pixel data in the intermediate image $\widehat{I_{n-k}}$ is so performed:

- pixels belonging to non-critical zones are assigned a weighted linear combination of data extracted from I_n and I_{n-M};
- pixels belonging to uncovered zones are assigned pixel data taken from image I_n only;
- pixels belonging to covered zones are assigned pixel data taken from image I_{n-M} only.

4 SOME EXPERIMENTAL RESULTS

The simulations were carried out on time-subsampled versions of the *Flower Garden*, *Table Tennis* and *Foreman* sequences in CIF format. The proposed motion estimation algorithm was compared with the block-based one, both to predict an images (belonging to a sequence) from the previous one and to interpolate missing frames in an image sequence. The results showed an appreciable improvement in the quality of the estimated motion fields (see figures 2, 3, 4, 5) and of the reconstructed images when using the proposed algorithm (see fig. 6). Current research is oriented at introducing object time-tracking techniques, in order to enhance the segmentation quality. Also, we are working on the reduction of the computational complexity, to make a cost-effective implementation of the algorithm possible.

Figure 2: Original current image (frame 26).

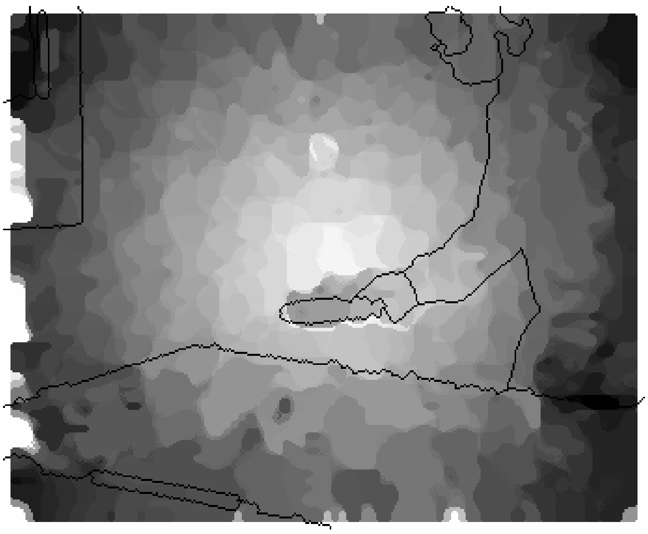

Figure 3: Magnitude of initial motion field block based estimate (luminance edges overimposed).

Figure 4: Vectors discarded during the robust regression phase (marked in black and overimposed to original frame).

Figure 5: Magnitude of final affine-regularized motion field (luminance edges overimposed).

Figure 6: Interpolated image (corresponds to skipped frame 25).

References

[1] C. Cafforio, F. Rocca, S. Tubaro, *Motion Compensated Image Interpolation.* IEEE Transactions On Communication, vol. 38, n. 2, February 1990, pp. 215-222.

[2] L. Vincent and P. Soille, Watersheds in Digital Spaces, *An Efficient Algorithm based on Immersion Simulations.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(6), Jun 1991, pp. 583-598.

[3] P.J. Rousseuw, F.R. Hampel, E.M. Ronchetti and W.A. Stahel, *The approach based on the influence function.* John Wiley and Sons, 1986.

[4] P. Migliorati, S. Tubaro, *Multistage Motion Estimation for Image Interpolation.* EURASIP Signal Processing: Image Communications, 1995, n. 7, pp. 187-199.

[5] Bierling Thoma, *Motion compensating interpolation considering covered and uncovered background.* Image Comunication, pag 191-192, May 1989.

[6] G. de Haan, P.W.A.C. Biezen. *Sub-pixel motion estimation with 3-D recursive search block-matching.* Signal Processing: Image Communication 6 (1994), pp. 229-239.

[7] F. Moscheni. *Spatio-Temporal segmentation and object tracking: an application to second generation video coding.* PhD thesis, Swiss Federal Institute of Technology, Lausanne, CH, February 1997.

[8] Y. Bar-Shalom and T.E. Fortmann. *Tracking and Data Association.* Academic Press, Inc., 1998.