NON-LINEAR TECHNIQUES IN VIDEO: THE MAVI COLLABORATION

Luc Van Eycken and all MAVI partners

Katholieke Universiteit Leuven, ESAT/MI2
Kardinaal Mercierlaan 94, B–3001 Heverlee, Belgium
Luc.VanEycken@esat.kuleuven.ac.be

ABSTRACT

In this paper, we'll describe the results of the Human Capital & Mobility cooperation network "Model based Analysis of Video Information" (MAVI). This network focused its cooperation on non-linear image processing, feature extraction and image sequence segmentation, and model based video compression techniques. To complement the results, shown in this paper, a list of joint publications with more scientific details is also provided.

1. THE MAVI PROJECT

The "Human Capital & Mobility" (HCM) program for scientific and technical cooperation networks sponsors the training and exchange of researchers between European countries, around a common topic. In our project, called "*Model based Analysis of Video Information*" (MAVI), the goal was to study and develop techniques for one of the more complex image processing applications: the compact representation of image sequences. Supported by the first ideas about MPEG4, which became known at the time of our project definition, we restricted the project to the integration of expertise on the different aspects of model based analysis techniques. The collaboration was focused on the following topics: 1) the extension of non-linear techniques to higher dimensions, including colour and motion, as a set of basic tools for further video image processing; 2) feature extraction and image (sequence) segmentation; 3) model based video compression techniques, especially for low bitrate applications.

This project of 2.5 years (including its extension) started in November 1994. Nine European partners participated: University of Leuven, Belgium (coordinator); University of Gent, Belgium; Tampere University of Technology, Finland; CMM, École des Mines de Paris, France; Aristotle University of Thessaloniki, Greece; CSTV, Consiglio Nazionale delle Richerche, Italy; University of Padova, Italy; Polytechnic University of Catalonia, Spain; École Polytechnique Fédérale de Lausanne, Switzerland.

Since the goal of the project is cooperation through exchange and training of people, the outcome of this collaboration is not some product, prototype or demonstrator. But of course, through the exchange and collaboration of researchers, some joint scientific results were obtained, often in collaboration with other projects of the partners. An overview of the research will be given in section 2.

The training aspect of the project was accomplished, not only by several locally organised courses, but also by two international activities. First of all, the project participated in a computer vision course, which was broadcasted by satellite, but which is also available on video tape. Furthermore, a MAVI course was organised in Finland, where tutorials were given by several partners. During the course, an audio conferencing link was also established to Greece.

2. OVERVIEW OF THE MAVI RESEARCH

The research covered by this project is subdivided into three work packages, according to the three topics mentioned in section 1, each described in one of the following sections. For each work package some research topics using non-linear techniques in video are described in more detail.

The results mentioned below describe of course only a part of the large amount of research that was carried out by all the partners on these topics. But the results described here can give you an idea about the possibilities of such a collaboration. More details about these and other results can be found in the references in section 3.

2.1. WP1: Non-linear video signal filtering

Video signals are usually corrupted by various types of noise starting from the formation process to the display mode. As a result, these signals are often preprocessed, filtered, prior to further analysis, e.g. segmentation. Real video signals contain an enormous amount of details and texture and are often corrupted by non-Gaussian, and sometimes non-additive, noise appearing in severe cases as im-

This work was carried out in the framework of the Community Research Programme with financial contribution by the European Commission.

pulsive noise. Furthermore the extra dimensions of colour and time have to be taken into account.

In addition to those described below, some of the research topics in this work package were: algorithm for detecting edges and lines; Volterra filtering; non-linear techniques for feature tracking; error concealment for MPEG2 [29].

Connected operators

Connected operators [22, 23] are non-linear morphological tools that allow us to simplify an image or a sequence of images while preserving the contour information. These filtering tools are extremely useful for analysis and segmentation. A set of criteria have been developed: size, contrast, complexity, motion and depth (distance from the camera). For non-increasing criteria a new formulation of the problem was needed resulting in a solution involving a dynamic optimisation procedure. Furthermore intensive work has been done on the efficient implementation of the operators.

Image enlargements and colour compression

The algorithm designed here, called Morphological Image Enlargement [1], reconstructs the image edges from a blurred enlarged image. In order to produce non blurred edges, a morphological image reconstruction is performed in the high contrasted regions. These regions are emphasised by computing the morphological gradient of the enlarged blurred image. The blurred information is then ignored at the edges and replaced by the reconstructed image. This technique is similar to the watershed segmentation transformation. The difference comes from the fact that no watershed lines and no separated regions are built.

Colour compression schemes can also be designed with the help of these techniques. Contrary to the classical colour compression algorithms which do not take into account the position of the colour pixels in the image, a morphology-based compression do not affect the same weight to the pixels if they fall in homogeneous regions or in edge zones.

Scale space for segmentation

Most edge detectors are based on partial derivatives of the image, and this makes the finding of edges extremely noise sensitive. Therefore, before the derivative is taken the image is filtered to improve its quality and reduce the corrupting noise. The width of the filter is not a priori known. Therefore, filters of increasing width are used. All filtered versions taken together are called the scale space. Edges that remain over several scales are interpreted as true edges.

In this project, the scale space based on morphological openings and closings was studied. Results of this study have been used in the motion estimation and parallax estimation research [7, 8].

Colour image filtering

A new family of multichannel L-filters that are based on the R-ordering principle were developed [17]. The coefficients of the filters can be optimised for a specific noise distribution with respect to the mean square error between the filter output and the desired, noise-free signal. Experiments with noisy colour images have shown that a very good noise reduction is achieved. The new family of filters has also been applied to motion vectors fields.

The evaluation of the optimal coefficients for the proposed multichannel L-filters requires a priori information on the signal statistics. To overcome this limitation, adaptive multichannel L-filters that are based on a variant of the LMS algorithm were introduced. Experiments involving colour images proved that the performance of the adaptive filters is very close to the performance of the corresponding optimal filters.

2.2. WP2: Segmentation and feature extraction

After the preparation of the signal by the filtering techniques, the next step in model based techniques is to analyse the signal and extract the relevant features. These features are used to segment the image in coherent parts. In model based techniques, feature extraction and segmentation are often interwoven, so they are considered here to be one task.

In addition to those described below, some of the research topics in this work package were: non-linear colour image restoration, colour image processing with B-spline modelling, motion based segmentation, segmentation based on morphological tools, segmentation based on texture analysis, automatic marker extraction for marker based segmentation, and object tracking.

Watershed segmentation

A new segmentation algorithm related to watershed has been developed. Unlike watershed which follows the principle of flooding, the new algorithm is based on rain-falling simulation. the rule of assigning labels can be derived from physics: a particle in free fall on a topographic surface will move, due to gravity, downward to the deepest neighbouring location. Besides the fast running time, the new algorithm is much more amenable to parallelisation than those based on flooding. The results are summarised in [6].

Furthermore, a new algorithm for computing the 3D watershed transformation is developed. This algorithm can easily be extended to different grids and connectivities. Unlike a 2D algorithm, the 3D algorithm is capable of separating objects in a scene which are included in several consecutive images.

Spatio-temporal segmentation

In this technique [16], the segmentation is expressed as a relaxation problem based on a Markov Random Field modelling and a Bayesian criterion. With this approach the spatio-temporal segmentation and the motion are simultaneously estimated. In order to handle efficiently camera motion (zoom and pan), the latter is first removed through global motion estimation by a frame matching technique. The image sequence is then pre-filtered to make it easier to segment. The subsequent spatio-temporal segmentation process takes into account the information from space and motion by departing from a static segmentation for which the local motion estimation is performed. The local object-based motion estimation relies on a parametric affine motion model and is performed by a matching algorithm using a robust estimator. Regions corresponding to a failure of the motion estimation are further split by a clustering on the luminance. Conversely regions characterised by a similar motion are merged by clustering in the motion parameter space. By exploiting the static segmentation, this algorithm assures very precise motion boundaries. In particular, it avoids problems related to occlusion and uncovered background.

Morphological moving object segmentation and tracking

As a non-linear approach to signal processing, multivalued morphology can deal with various images such as colour images, as well as motion vector fields. Using the set-theoretical methodology for image sequence analysis, multivalued morphology can estimate many features of the geometrical structure in the signals, where the partition of motion is just a particular case. This way, generic motion segmentation can be achieved if multiple sources, e.g., colour and motion, are taken into account simultaneously. Using the strategy of divide-and-conquer, a big and difficult problem can be solved by dividing it into smaller and easier ones. Based on the above arguments, the proposed algorithm was developed in which the result of a hierarchical motion segmentation was combined with the results of a spatial segmentation in order to provide a precise partition of the motion field. Furthermore, the problem of motion tracking was solved at the same time in order to avoid the random fluctuation of the motion segmentation and establish links of moving objects in successive frames. The algorithm was presented in [10].

Face segmentation and tracking

A face localisation module has been developed that locates skin-like regions by performing colour segmentation in HSV colour space [24, 25]. The hue and saturation domain, which describes the human skin colour, can be defined or estimated a priori and used subsequently as reference for any skin colour. The oval shape of a face can be approximated by an ellipse. In a first step, connected components are determined by applying a region growing algorithm at a coarse resolution of the segmented image. Then, for each connected component with a given minimum size, the best-fit ellipse is computed on the basis of moments.

The proposed approach for facial feature extraction is based on the observation that, in intensity images, eyes and mouth differ from the rest of the face because of their low brightness. Therefore, in the following, the intensity information in the interior of the connected components derived in face localisation is employed. The first approach used a modified watershed method and determines the eye and mouth regions by flooding the grey-level relief. The second approach is based on min-max analysis and evaluates directly the x- and y- projections of the grey-level relief [26]. In a preprocessing step, dark regions in the interior of the connected components are enhanced by using a grey-scale erosion. Fuzzy rules have been employed to access how well a pair of minima meets the requirements for eye candidates. Additional fuzzy rules have been developed for mouth candidates.

Tracking is performed by using an active contour model [27]. The contour of the segmented face is taken as initial snake. By employing regions instead of edges a fast processing and an increase in robustness for the face recognition process are ensured. Snaxels are selected based on equal Euclidean distances. The exterior forces of the snake are defined based on colour features. The dynamic behaviour of the snake is controlled by a Greedy algorithm and it yields a local minimum of the energy.

2.3. WP3: Low bit-rate coding of video signals

One of the objectives of this project is to explore model based video coding as a way to achieve these low bit-rates. The video information will be decomposed into different objects according to some selected models. Although these objects can correspond to real objects, this is not a requirement. The image sequences will be coded by the values of the model parameters, including the motion of the objects.

Dynamic coding

Dynamic coding [18] is based on a content representation in which different objects in the scene may be encoded by different coding techniques chosen to be the most appropriate ones for their characteristics and/or coding parameters. This allows to allocate different parameters to different regions depending on their importance and specific features, such as quantisation steps, size of coded blocks, targeted bit-rate, adapted coding algorithms, motion search window, and so on.

This technique was presented at the October subjective tests for MPEG4 and turned to out to be the best European object based technique for very low bit-rate coding of video allowing content interactivity.

Combined motion estimation and segmentation

An important constraint is the limited number of bits available for segmentation and texture information combined. Therefore, careful examination of where segmentation information is vital and where it can be ignored is necessary. We chose motion based segmentation, since this normally yields less regions, certainly in head-and-shoulder sequences. To get even better results, we try to combine the motion estimation step and the segmentation, using a cost function giving a reliability of the motion information based on spatial and temporal continuity. The motion is described with affine motion vectors.

The obtained segmentation is not a hard segmentation, but it gives a probability for each pixel to belong to a certain segment. This allows us to select the most useful object border, taking into account the limited number of bits available for the segmentation coding. Superquadrics, which are generalised ellipses/rectangles, are used as curves for describing the object border. With a predefined set of quantised parameters many simple objects can be approximated, taking into account the above-mentioned probabilities. If the object is more complex and the bit-rate allows for it, additional superquadrics are added or subtracted. The implementation is described in [30].

Coding using 3D-objects

In this coder for head and shoulder video sequences [2], regions are modelled as projections of 3D rigid objects onto the camera image plane. Model failure regions are associated with the eyes and mouth, which are coded and treated differently. The motion of each rigid body of the scene can be described by a non-linear dynamic system whose state consists of the motion parameters and the scaled depths of the feature points. The state of the dynamic system is estimated from the measurements of the positions of the feature points on the image plane by means of an extended Kalman filter. Therefore, at each step, the estimator takes the positions of the features on two consecutive frames as input and yields the estimated state. The estimated state describes the motion and the shape of the model compliance objects, and therefore it is possible to synthesise these objects using motion and shape information and the texture of the first image coded in intra mode.

3. REFERENCES

[1] A. Albiol, J. Serra, "Morphological image enlargements", *CMM report, Paris School of Mines, N-51/96/MM*, Dec. 1996.

[2] G. Calvagno, L. Celeghin, R. Rinaldo, L. Sbaiz, "Statistical Based Motion Estimation for Video Coding", in *Proceedings IEEE 1996 International Conference on Image Processing*, pp. 105–108.

[3] J. Crespo, F. Meyer, "Morphological segmentatioon using texture and coding cost," in *Proceedings of the IEEE Workshop on Nonlinear Signal and Image Processing*, Neos Marmaras, Halkidiki, Greece, June 1995.

[4] G.M. Cortelazzo, G.A. Mian, R. Pazolari, A. Chimienti, "Nonlinear Filtering Methods for 2-D Signals and Their Application to Granular Noise Reduction", in *Proceedings of COMCON 5*, Rhodos, Greece, June 1995.

[5] G.M. Cortelazzo, F. Pollastri, M.L. Rastello, "Colorimetric evaluation of media for recording works of art", in *Proceedings of Science and Technology for the Safeguard of Cultural Heritage in the Mediterranean*, Basin, Catania, Italy, Nov. 1995.

[6] B. Cramariuc, A. Moga, M. Gabbouj, "Image segmentation by component labelling," in *Proceedings of the International Conference on Digital Signal Processing*, Cyprus, June 1995.

[7] P. De Smet, D. De Vleeschauwer, M. Gabbouj, "Improving the Accuracy of Motion Vectors obtained by Block Matching", in *Proceedings NORSIG'96*, Espoo (Finland), Sep. 1996.

[8] D. De Vleeschauwer, P. De Smet, M. Gabbouj, "Improvement of the Block Matched Parallax Field by Non-linear Diffusion", in *Proceedings NORSIG'96*, Espoo (Finland), Sep. 1996.

[9] C. Doignon, G.M. Cortelazzo, R. Frezza, "A robust camera calibration technique for high resolution image acquisition", in *Proceedings of the European Conference on Multimedia Applications, Services and Techniques*, Louvain-La-Neuve, Belgium, May 1996.

[10] C. Gu, T. Ebrahimi, M. Kunt, "Morphological spatio-temporal segmentation for content-based video coding", in *International Workshop on Coding Techniques for Very Low Bit-rate Video*, Tokyo, Nov. 1995.

[11] C. Jacquelin, D. De Vleeschauwer, "Segmentation of Texture Image Sequences", in *Proceedings NOR-SIG'96*, Espoo (Finland), Sep. 1996.

[12] F. Marqués, S. Fioravanti, P. Brigger, "Coding of image partitions by morphological skeletons using overlapping structuring elements", in *Proceedings of the IEEE Workshop on Nonlinear Signal and Image Processing*, pp. 250–253, Halkidiki, Greece, June 1995.

[13] F. Marquez, B. Marcotegui et al., "A segmentation-based coding system allowing manipulation of objects (Sesame)", in *Proceedings of the International Conference on Image Processing*, Lausanne, Sep. 1996.

[14] F. Marquez, B. Marcotegui, F. Meyer, "Tracking areas of interest for content-based functionalities in segmentation-based video coding", in *Proceedings ICASSP'96*, Atlanta (USA).

[15] F. Meyer, O. Ribes, "Contour coding system in the hexagonal raster," in *Proceedings of the IEEE Workshop on Nonlinear Signal and Image Processing*, Halkidiki, Greece, June 1995.

[16] F. Moscheni, *Spatio-temporal segmentation and object tracking: An application to second generation video coding*, PhD thesis, École Polytechnique Fédérale de Lausanne, Switzerland, 1997.

[17] N. Nikolaidis and I. Pitas, "Multichannel *L* Filters Based on Reduced Ordering", *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 6, no. 5, pp. 470–482, Oct. 1996.

[18] E. Reusens, R. Castagno, C. Le Buhan, L. Piron, T. Ebrahimi, M. Kunt, "Dynamic coding — An overview", in *IEEE International Conference on Image Processing*, Vol. II, pp. 377–380, Lausanne, Switzerland, Sept. 1996.

[19] P. Salembier, P. Brigger, J.R. Casas, and M. Pardàs, "Morphological operators for image and video compression", *IEEE Transactions on Image Processing*, 5(6):881–898, June 1996.

[20] P. Salembier, F. Marqués, M. Pardàs, R. Morros, I. Corset, S. Jeannin, L. Bouchard, F. Meyer, B. Marcotegui, "Segmentation-based video coding system allowing the manipulation of objects", *IEEE Trans. on Circuits and Systems for Video Technology*, 7(1):60–74, Feb. 1997.

[21] P. Salembier, F. Meyer, P. Brigger, L. Bouchard, "Morphological operators for very low bit rate video coding", invited paper, in *Proceedings of IEEE International Conference on Image Processing, ICIP'96*, vol. III, pp. 659–662, Lausanne, Switzerland, Sep. 1996.

[22] P. Salembier and J. Serra, "Flat zones filtering, connected operators and filters by reconstruction", *IEEE Transactions on Image Processing*, 3(8):1153–1160, Aug. 1995.

[23] P. Salembier, L. Torres, F. Meyer, C. Gu, "Region-based video coding using mathematical morphology", invited paper, in *Proceedings of IEEE*, 83(6):843–857, June 1995.

[24] K. Sobottka and I. Pitas, "Localization of facial regions and features in color images," in *Proceedings of the 4th Open Russian-German Workshop on Pattern Recognition and Image Analysis*, pp. 134–138, Valday, The Russian Federation, March 1996.

[25] K. Sobottka and I. Pitas, "Extraction of facial regions and features using color and shape information," in *Proceedings of the 12th International Conference on Pattern Recognition (ICPR 96)*, vol. III, pp. C421–C425, Vienna, Austria, Aug. 1996.

[26] K. Sobottka and I. Pitas, "Looking for faces and facial features in color images," *Journal of Pattern Recognition and Image Analysis: Advances in Mathematical Theory and Applications*, Russian Academy of Sciences, 1996.

[27] K. Sobottka and I. Pitas, "Segmentation and tracking of faces in color images," in *Proceedings of the 2nd International Workshop on Automatic Face and Gesture Recognition*, Killington, Vermont, U.S.A, Oct. 1996.

[28] L. Torres and M. Kunt, editors, *Video Coding: The Second Generation Approach*. Kluwer, 1996. ISBN: 0 7923 9680 4.

[29] S. Tsekeridou, I. Pitas and C. Le Buhan, "An Error Concealment Scheme for MPEG-2 Coded Video Sequences", technical report, *submitted to ISCAS 97*, Oct. 1996.

[30] T. Wuyts, L. Van Eycken, A. Oosterlinck, "Segmentation based coding for very low bitrates", in *Visual Communications and Image Processing '97*, vol. 3024, part II of Proceedings of the S.P.I.E., pp. 690–698, San Jose, California, USA, Feb. 1997.