UNIVERSITAT POLITÈCNICA DE CATALUNYA

Departament de Teoria del senyal i comunicacions

MULTIRESOLUTION IMAGE SEGMENTATION BASED ON COMPOUND RANDOM FIELDS: APPLICATION TO IMAGE CODING

Autor: Ferran Marqués
Director: Antoni Gasull

Barcelona, Diciembre de 1992

CHAPTER VII

CONCLUSIONS

The goal of this chapter is twofold. Its first section is devoted to the presentation of a summary of the developments carried out in this work. In this way, the different parts of the work are briefly summed up and commented. On its turn, the second section deals with the possible extensions of this work. Extensions deal with the segmentation technique as well as with its coding application. In this way, a method for segmenting image sequence is outlined. Furthermore, some possible improvements on the coding scheme presented in the previous chapter are pointed out. Some of these extensions are currently under development while the study of some others have not started yet.

VII.1.- Summary of developments

The problem of segmenting gray level, still images has been addressed in this work. Segmentation has been defined as a process that divides completely an image into a set of homogeneous, connected regions, related to the objects in the scene. The definition of a general purpose segmentation technique has been revealed as being a rather complicated task. This complication is owing to the huge amount of different kinds of data that a segmentation technique may have to handle. Nevertheless, the study of classical techniques has shown that the use of some specific tools results in a big step towards this goal. Among these tools, elaborated image models and multiresolution image analysis lead to a clear improvement on the performance of segmentation algorithms. Therefore, the proposed segmentation method relies on both techniques.

Stochastic image models have been analysed, as well as their application to the problem of image segmentation. For this purpose, Compound Random Fields (CRFs) are particularly suitable. CRFs allow the description of images in terms of both boundary and texture information. Furthermore, assuming that images are modelled by CRFs, the segmentation task can be formulated as a MAP estimation. This kind of algorithms achieves high quality segmentation results. However, these segmentation approaches usually require stochastic maximisation techniques.

When using stochastic relaxation approaches, the optimum solution is ensured to be reached. Nevertheless, this kind of methods are computationally very demanding. On the other hand, although deterministic relaxation approaches may get trapped in local maxima or minima, they may achieve good quality results in a feasible amount of time. The quality of the obtained results mainly depends on the initial conditions of the algorithm. Therefore, in this work, a deterministic approach has been chosen, having in mind the importance of setting the correct initial conditions. This approach performs the maximisation by carrying out local, iterative boundary refinements.

A second problem related to the use of CRF models in segmentation is the estimation of the model parameters. The estimation of elaborated image model parameters is a difficult task which requires time consuming techniques and may result in estimationsof low reliability. The image model used in this work is a CRF formed by a set of white Gaussian random fields in its upper level and by a Strauss process in its lower level. The reason for choosing white GRFs for the upper level is that this kind of model, when used in a CRF, leads to a good trade-off between capacity of characterising regions and computational load when estimating its parameters. On its turn, the lower level is modelled by a Strauss process, given its good performance when modelling labelled images. Furthermore, Strauss processes can also be utilised for modelling the behaviour of contour images.

A monoresolution segmentation technique has been proposed relying on the joint use of the deterministic maximisation approach and the image model above commented. This segmentation technique leads to good results which always improve the quality of the initial segmentations. The influence of the low level parameters in the final result has been studied. This study has shown that similar results are achieved for a large range of parameter values. Moreover, in all cases, segmentations are carried out in a feasible amount of time. Although the method yields good results, it presents some drawbacks. These drawbacks are mainly produced by the fact that, in order to reduce its computational load, the algorithm carries out a maximisation procedure in a local, deterministic way. Therefore, the quality of the initial segmentation is critical to the quality of the final result.

A novel segmentation algorithm relying on the joint use of an image model based on CRFs and a multiresolution decomposition has been presented, in order to overcome the above problems. A Gaussian pyramid has been chosen as multiple resolution decomposition. The reasons for choosing this kind of decomposition are mainly related to the data simplification that it performs. That is, textures present at low levels are gathered, at high levels, into a few pixels whose statistics are more Gaussian-like than those of original ones.

This multiple resolution segmentation technique relies on the previous monoresolution method. That is, the image is first decomposed into a set of images at different resolutions. At each level of the decomposition and starting by the coarsest resolution, images are segmented using the monoresolution algorithm. The final segmentation at each level is utilised as initial segmentation for the next finer level of the decomposition. This procedure is performed through the whole decomposition down to the finest resolution level, which is the original image itself.

With such an approach, the negative effects of the deterministic maximisation procedure are reduced. This improvement is owing to the fact that the algorithm combines the information of the different resolutions in order to segment the image. In addition, since every segmentation is performed assuming the same kind of image model, initial segmentations at each level are more consistent with the expected final segmentation. Finally, local maxima are easier to avoid when using a multiresolution approach. Solutions achieved at coarse resolution lead finer resolution steps to initial points which are closer to the actual optimum. Therefore, the performance of the method increases with respect to the monoresolution algorithm, specially when dealing with textured areas.

In spite of this improvement, the basic multiresolution algorithm still presents some drawbacks. These drawbacks are mainly related to the use of a constant image model at every decomposition level and to the inability of the algorithm for detecting interior regions. The first problem has been solved by adapting the model parameters to the data present at each level of the decomposition. To this goal, an unsupervised estimation of the model parameters relying on the information contained in the Laplacian pyramid has been performed. This unsupervised segmentation method yields good performance both in computational load and segmentation quality. Regarding the computational load, the multiresolution approach require less computation than the monoresolution one, for same quality results.

The problem of detecting interior region has been faced from the point of view of discriminating between flat, contrasted areas and dense, contrasted fluctuation. Morphological tools have been applied, given that they are known to efficiently perform such kind of discrimination. In this way, the residue with the morphological centre has been used as basic transform. The centre is computed from the open_close, the close_open and the identity operator. This transform has been applied, at each level, on the difference between the original image and the mosaic representation of the segmentation. After a cleaning step, the elements resulting of performing this transform have been used as seeds for the possible creation of new regions. This segmentation scheme leads to high performance results both in quality and computational load. This method yields final segmentations nearly containing all the details in the original image, while does not spoil the segmentation on textured areas. In addition, the average computational load for segmenting images of 256x256 pixels is of 27 seconds in a Sun Sparc II workstation.

Finally, this segmentation technique has been applied to the task of image coding. Segmentations yield two different kinds of information to be coded: texture and boundary information. Boundary information coding requires much more effort than texture coding. Hence, the problem of coding contour images produced by a previous segmentation of the scene has been addressed. Contour images are defined on an hexagonal lattice for allowing one-to-one relationship between contour representations and segmentations.

When coding contour images, two different information have to be coded: shape and location information. The proposed coding scheme relies on derivative chain code techniques for coding region shapes. The information regarding region locations is introduced in the chain code itself. This is done by making use of the concept of triple point. Symbols within chains are coded by using Huffman techniques. Texture information is coded by using one byte per region (texture characterisation by its mean value). This coding scheme leads to compression ratios of 25 in average.

VII.2.- Current and future research lines

Three different lines of research are proposed in this section. The first one deals with the extension of the proposed still image segmentation method to an image sequence segmentation technique. This extension involves the definition of a stochastic image sequence model. Assuming the image model proposed in this work as basis, CRFs have to be defined in a three dimensional domain. This new definition leads to the extension of the concept of clique to a three dimensional space. Note that the assumption of isotropy made when defining potentials on the two dimensional space cannot be directly made in this case. A clique involving sites belonging to two different frames may not represent the same information as a clique totally contained in a single image.

Nevertheless, first experiments obtained by making the assumption of isotropy yield good segmentation results. An example of this technique can be seen in Figure VII.1, where the segmentation of a sequence of four frames is shown. However, in some cases, this assumption leads to segmentations presenting volumes (three dimensional segments) with long, thin elongations. Currently, the correct relation between the different clique potentials is being analysed in order to improve the segmentation performance. This analysis is being performed as in the two dimension case; that is, by isolating the influence of each parameter in the whole model. In the image sequence case, the potential function characterising the Gibbs distribution can be expressed as:

$$U(x) = \frac{1}{T}[\, k_{1s}V_{1s} + k_{2s}V_{2s} + k_{1t}V_{1t} + k_{2t}V_{2t} \,] \quad , \qquad \text{(VII.1)}$$

where the notation follows the definition in Chapter III and the subindices s and t denote the space and time domains, respectively. The set of parameters characterising the model can be reduced to four:

$$U(x) = \frac{V_{1s}}{T}\left[k_{1s} + k_{2s}\frac{V_{2s}}{V_{1s}} + \frac{V_{1t}}{V_{1s}}\left(k_{1t} + k_{2t}\frac{V_{2t}}{V_{1s}\,V_{1t}} \right)\right] =$$

$$= \frac{1}{T^*}\left[k_{1s} + k_{2s}V_s^* + R_{st}^*\left(k_{1t} + k_{2t}V_t^* \right)\right] , \qquad \text{(VII.2)}$$

where T^* controls the relative importance between the lower level and upper level models, V_s^* characterises the spatial boundary model, R_{st}^* sets the relation between the spatial and temporal parts of the lower level model and V_t^* characterises the temporal boundary model.

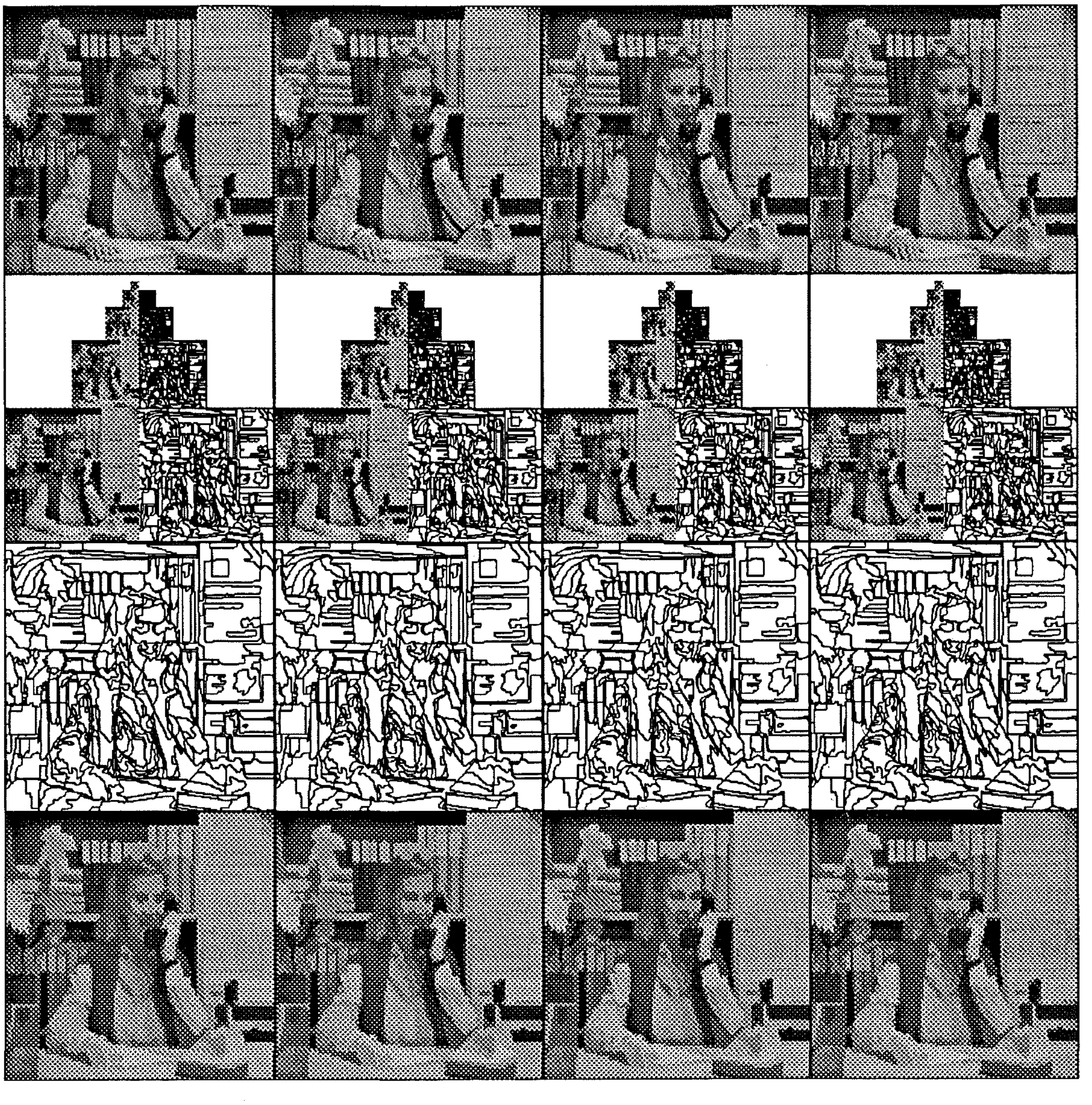

Fig. VII.1.- Example of image sequence segmentation

The second research line copes with the improvement of the contour coding scheme. The shape coding technique presented in this work takes advantage of some global constraints in the contour shape. A way for further improving this method is by exploiting local contour constraints and not only general ones. These local constraints are mainly introduced by the algorithm that selects each initial point of the chains as well as the contour tracking algorithm. Initial points are selected by a normal scanning (from top to bottom and from left to right). Therefore, after a initial point, only *s* and *r* movements are allowed. Due to this constraint, first movements in a chain can be coded with just one bit until the first *r* movement is made.

On its turn, the contour tracking procedure follows a clockwise direction. Therefore, when being in a lattice site, if a left turn (*l*) is selected, neither a right turn (*r*) nor a straight ahead (*s*) movement were possible. The two sites related with the withdrawn movements (*r* and *s*) can be marked in the lattice as forbidden sites. When choosing an *s* movement, an analogous situation appears; that is, the site at the right can be marked as forbidden. Due to these marks, future chains are constrained and codes may not have three possibilities. In these cases, when the number of possible codes is two, one bit can be used for coding movements. Moreover, zero bits can be used when the contour only has a possible path to follow. Note that, in order to marked forbidden sites, just past and actual information is used. Therefore, this task can be performed in the coder as well as in the decoder. By using this new scheme, figures of 1.12 bits per pixel of contour (in average) are achieved for shape coding. Current work deals with introducing Huffman techniques, as well as triple points location coding, in this new scheme.

The third research line copes with the coding of the texture information. As pointed out in Chapter VI, textured regions cannot be represented by their mean value when aiming at coding purposes. Therefore, more elaborated coding schemes have to be used in such cases. In this way, a classification of regions yielded by the segmentation should be performed in order to discriminate between textured and non-textured regions. On the first class, a textured coding scheme involving transform coding will be applied, whereas, for the second class the current scheme is sufficient.