MULTIRESOLUTION IMAGE MOTION DETECTION AND DISPLACEMENT ESTIMATION

Shuwu Song * , Mengyang Liao and Jiamei Qing

Radio & Information Engineering Department
Wuhan University, Wuhan, PR China

ABSTRACT

In our paper, we developed a motion vision system in which we can detect the moving object and estimate the displacement based on the human visual characteristics and multiresolution image or pyramid. The system consists of the following four parts.

1. Using the temporal gradient, AND and dynamic thresholding operations to obtain the primary mask.
2. Region growing.
3. Hierarchical object detection.
4. Displacement Estimation. Each frame of a motion sequence is broken into local region (edge) and a search undertaken to discover how the image pattern within a given region appears displaced. This search takes the form of the motion channels. Then we can obtain the estimation of displacement from the outputs of all motion channels. The authors also proposed a correlative measure to match the patterns and the measure yields a good performance.

INTRODUCTION

Motion detection and displacement estimation is rapidly becoming a major area of computer vision and image processing. The upsurge in activity is based on the need for more effective interframe data compression as well as on the theoretical interest in the subject. Motion estimation in a sequence of images may have two different purposes: 1) Pattern recognition purpose where detection and displacement estimation of moving object is a tool for scene understanding; 2) Coding purpose, by taking into account the temporal redundancy present in a picture sequence, motion estimation is a tool for motion compensated prediction of intensity in a picture sequence. In [1] a hierarchical object detection algorithm was proposed based on multiresolution image structure by the authors. P.J.Burt et al brought about a multiresolution flowthrough motion analysis method [2] which is effective and can be easily implemented in special purpose hardware. In this paper, based on human visual characteristics and multiresolution image, the authors present a new motion detection and displacement estimation mothod as well as a new correlative measure which results in a more accurate estimation of displacement.

----

* Shuwu Song is present with the Dept. of Electronics, Zhongshan University, Guangzhou 510275, PR China

MULTIRESOLUTION IMAGE (PYRAMID)

Multiresolution image (pyramid) is a data structure within which the imput image is represented at successively reduced resolution [3]. The motivition for computing a structure description is to spend a fixed computational cost to transform the information in each image into a representation in which searching and matching are more efficient. Many researchers have shown that the efficiency of searching and matching process can be dramatically improved by performing the search at multiple resolution. Many basic image operation can also be performed efficiently within pyramid structure [4-9]. P.J.Burt defined four pyramid-based "computational tools": Gaussian pyramid, Recursive interpolation, Laplacian pyramid and Hierarrchical Descrete Correlation (HDC) [10-12].

A general statement of the local correlation between images is:

$$C(i,d) = A_1[F[f(i)].F[g(i+d)]]$$

here correlation is computed between a local pattern in f centered at point i and a pattern in g centered at i+d , F[ ] is a operator or filter which is used to preprocess the image before correlation is computed. The operator . may be any which compute pixel comparisions. A_1[ ] is an average value operator and the HDC process is defined through A_1[ ]. C is a measure of the similarity or match confidence between the two local patterns. To find the portion of g which matchs the pattern centered at i we look for that d which yields the largest value of this measure. The performance depends on the image content, shape and size of window as well as the degradation and noise in the images. P.J.Burt has studied five correlative measure [12]. Taking the performence and computational cost into account, we propose the following correlative measure:

$$C(i,d) = \frac{A_1[L_{1k}(i).L_{2k}(i+d)]}{(A_1[L_{1k}^2(i)])^{1/2}(A_1[L_{2k}^2(i+d)])^{1/2}}$$

where L_{1k} and L_{2k} are the kth level of two laplacian pyramids. We use this correlative measure to form the simple motion channels in each spatial frequency band of the images. From the experimental results we can see that the proposed measure has a better performence. The computational cost lies between the varince normalized correlation and Laplacian filtered correlation.

MOTION DETECTION

First we use the frame-to-frame differences to derive the primary mask of the moving objects. The difference pictures between previous and current frame as well as between current and next frame are threholded and logically AND-ed. The primary mask will not completely cover the object in the input frame if it has a fairly homegenious gray level and does not completely displace itself from one frame to the next. It is not possible to lower the threshold too much. Then the dynamic threshold proposed by the authors[13] is applied for the purpose of detection of primary mask of moving objects.

Secondly a region growing algorithm is applied on the primary mask. The 8-connected neighborhood of the investigated pixel is searched for pixels already marked by the primary mask. If marked neighbors exist then their average grayvalue is calculated and compared with the grayvalue of pixel in question. If their difference lies within a certain confidence interval the pixel is marked as part of the mask.

Finally, since the neural coding of images is in terms of edge elments, with orientation and contrast specified in the code for each element, the gray level in any uniform region of an image is not represented in the code at all but is perceptually inferred from the edge transition surrounding the region, we are only interested in the edge patterns of the moving objects, the authors proposed a hierachical object detection strategy which is supported by the image Gaussian pyramid and gradient magnitude and direction pyramids. The gradient direction information, whose computation is under the guidance of the planning, has been used through the agorithm. Such use makes the detection successful and fast. (For details please see [1].) Then we estimate the displacement of edge patterns of moving objects.

DISPLACEMENT ESTIMATION

In the process of motion detection, we obtain the image patterns (edge element), then we determine how the image appears displaced in the subsquent frames of image sequence. In effect, each frame of a motion sequence must be broken into local regions and a search undertaken to discover how the image pattern within a given region appears displaced in the next and subsquent image frames. This search takes the forms of matching process in which patterns are directly correlated. In our motion estimation scheme, we use the above proposed correlative measure to correlate the edge patterns to determine the displacement.

$F_1(i,j)$ and $F_2(i,j)$ are the two neighboring frames of image sequence, $L_{10}(i,j)$, $L_{11}(i,j)$, ..., $L_{1k}(i,j)$, ..., $L_{1N}(i,j)$, and $L_{20}(i,j)$, $L_{21}(i,j)$, ..., $L_{2k}(i,j)$, ..., $L_{2N}(i,j)$ are their Laplacian pyramids respectively. Since a shift of one sample at level k corresponds to a shift 2 pixels in the image, if the displacement $|d|<2^k$, the analysis would be performed at kth level of the pyramids. Like Burt's method, the analysis of motion in different direction and at different velocities may be performed in a set of more or less independent motion channels, each channel examine the image for evidence of motion in its particular direction. The channel obtainning the most compelling evidence at any given point in image is assumed to be the most nearly correct, while simple interpolation between channel outputs privides a more accurate estimate of actual displacement. To conserve computation, the number of channel is kept to a minimum. In our implementation, 9 simple channels are formed in each patial frequency band:

$$M_{\alpha,\beta}(i,j) = \frac{A_1[L_{1k}(i,j).L_{2k}(i+\alpha,j+\beta)]}{(A_1[L_{1k}^2(i,j)])^{1/2}(A_1[L_{2k}^2(i+\alpha,j+\beta)])^{1/2}}$$

where α and β are -1, 0, 1 respectively, forming zero, up, down, left, right and four diagonal channels. Zero, left and right channel outputs are

$$M_{Zk}(i,j) = M_{0,0}(i,j)$$

$$M_{Lk}(i,j) = M_{0,-1}(i,j)$$

$$M_{Rk}(i,j) = M_{0,1}(i,j)$$

The output of horizontal compound channel, displacement in horizontal direction and the corresponding confidence, is given by

$$D_{rtk}(i,j) = 2^{k-1}\frac{M_{Rk}(i,j) - M_{Lk}(i,j)}{M_{Zk}(i,j) - M_{Lk}(i,j)}$$

$$C = M_{Zk} - M_{Lk}$$

$$(\text{ While } M_{Rk} > M_{Lk})$$

$$D_{rtk}(i,j) = 2^{k-1}\frac{M_{Rk}(i,j) - M_{Lk}(i,j)}{M_{Zk}(i,j) - M_{Rk}(i,j)}$$

$$C = M_{Zk} - M_{Rk}$$

$$(\text{ While } M_{Rk} < M_{Lk})$$

Then in the same way we can obtain 6 other simple channel outputs. The outputs of these simple channels are combined in group of three to form the other three compound motion channels. Finally the outputs of four orientation channels are compared and combined into a single motion estimation for each edge pattern within the frequency band and the optical field of image patterns is obtainned.

EXPERIMENTAL RESULT

Figure 1 is Guassian pyramid of the human chromosome image, bottom level size of the pyramid is 256*256 while the top level size is 32*32. Figure 2 and Figure 3 are the optical fields of the chromosome obtained by using the Laplacian and the proposed correlative measures respectively when the image shifts [+1, -1] pixels. It can be seen that the proposed correlative measure has a better performance than the Laplacian correlative measure. Figure 4 is the optical field of the chromosome obtained by using the proposed measure when the image shifts [+3, -3] pixels and the estimation is made at the third level of the Laplacian pyramid. And the Figure 4 is the estimation result obtained at first level of the pyramid.

REFERENCES

[01] S.Song et al, Intelligent Object Detection in Multiresolution Environment, ASME, 1988, Shenzhen, China

[02] P.J.Burt et al Multiresolution Flow-Through Motion Analysis, Proceedings of Conference on Computer Visin,Pattern Recognition and Image Procesing, pp246-252, 1983

[03] P.J.Burt, The Pyramid as an efficient Computation Structure, in Multiresolution Image Processing and Analysis, Ed. by A.Rosenfeld, Springer-Verlag,Berlin, 1984

[04] S.L.Tanimoto and T.Pavlidis, A Hierachical Data Structure for Picture Processing, CGIP 4,1975, pp104-119

[05] M.D.Kelly, Edge Detection by Computer Using Planning, Machine Intelligence 6, 1971, pp397-409

[06] P.U.Sankar et al, A Gestalt-Guided Heuristic Boundary Follower for X-ray Image of Lung Nodules, IEEE Trans. PAM-4,3,pp320-331, 1982

[07] J.Sklansky et al, Tow-Resolution Detection of Lung Tumors in Chest Radiograph in Multiresolution Image Processing and Analysis, Ed. by A.Rosenfeld, New York, 1984

[08] L.G.Minor and J.Sklansky, Detection and Segmentation of Blobs in Infrared Image, IEEE Trans. SMC-11, pp194-201, 1981

[09] S.Tamura et al Plan-based Boundary Extraction and 3-d Reconstruction for Orthogonal 2-D Echocardiography, Pattern Recognition 20,2, pp155-162, 1987

[10] P.J.Burt, Fast Filter Transform for Image Processing, CGIP 16 pp20-51,1981

[11] P.J.Burt et al, The Laplacian Pyramid as a Compact Image Code, IEEE Trans. Com-31, 4, pp532-540, 1984

[12] P.J.Burt et al,Local Correlation Measures for Motion Analysis, A Comparative Study, Proceeding of Conference on Pattern Recognition and Image Processing, pp269-274, June 1982

[13] S.Song et al Multiresolution Image Dynamic Thresholding, in this Proceedings.

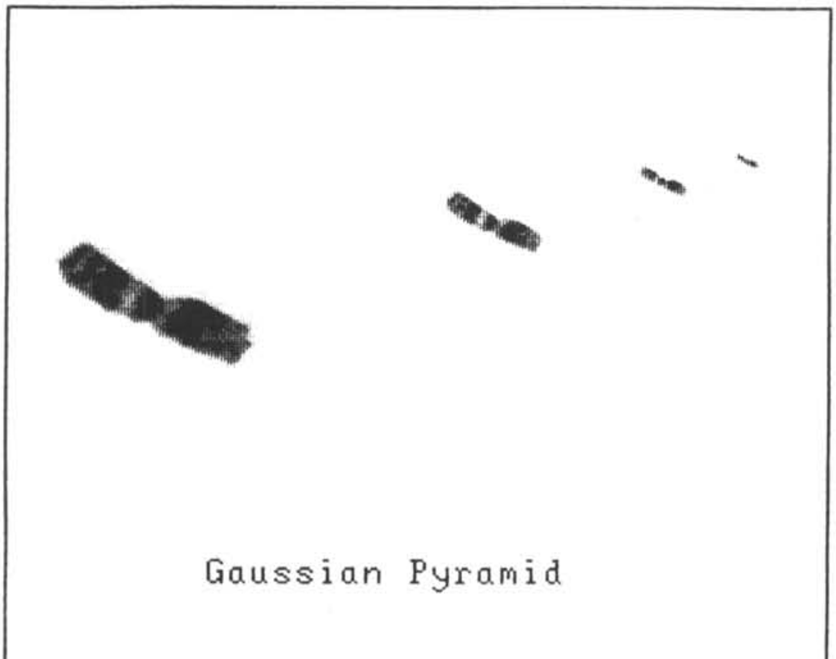

Fig. 1

Fig. 3

Fig. 2

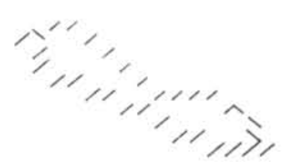

Fig. 4

Fig. 5