Proceedings of the 6th WSEAS International Conference on Robotics, Control and Manufacturing Technology, Hangzhou, China, April 16-18, 2006 (pp181-186)

A New Approach to Adaptively Segment Natural Scene Images for Target Localization for Real-Time Applications

M. SALMAN CHAUDHRY, JAVAID IQBAL AND AKHTAR NAWAZ MALIK
Departments of MTS and Computer Engineering
College of Electrical and Mechanical Engineering
Peshawar Road, Rawalpindi
PAKISTAN

Abstract: - This work formulates a new method to segment target areas from natural scene images that have a grey scale average different from that of the background. Most of the available methods suffer from high implementation complexity which makes them unsuitable for real-time applications such as robotics. The presented method is an enhancement to an existing method which can only segment bright targets. This is done by using a new graphical strategy to extract the background class using the analysis of image's wavelet subband PDF. The new strategy uses the center of masses of the accepted target threshold groups rather than using the globally maximum local minima which has dramatically improved the results. Moreover a novel validation scheme is introduced which simply eliminates the unwanted detections based on their characteristics. The method is applied to a variety of real world cases and is found to give exceptional results.

Key-Words: - Wavelets, Real-time, Subband, Class, Validation

1 Introduction

There exit many methods for segmenting natural scene images but they lack the ease of implementation for real-time applications such as robotics. This gives rise to the need to invent segmentation algorithms which are capable of generating accurate results in real-time. Among these techniques, the discrete wavelet transform is most widely used due to its efficient multi-resolution analysis characteristic. Unfortunately, the researchers have yet developed either very complex methods for accurate segmentation or very simple algorithms that are fast but not up to the task. Therefore a new method is designed that can give high quality results while at the same time being able to run in real-time using easy implementation based on hardwired techniques (e.g. VLSI).

Zhang [4] has presented a method to segment bright targets from non-uniform backgrounds using undecimated DWT which is a limited application. In addition there is no suitable technique to find/eliminate the background class. The presented method uses two separate levels of the 1-D wavelet transform of PDF of the image's wavelet approximation subband to extract the background class and local minima of the PDF. The final thresholds are found to give much better results.

We have enhanced the existing algorithm [4] in three areas; 1) a new approach to detect the background class, 2) a new method for detection of target threshold groups and their resultant, 3) a new segmentation validation scheme, thus yielding extraordinary results.

The following section briefly discusses the undecimated DWT (discrete wavelet transform), section 3 discusses the proposed algorithm and section 4 gives case studies of the proposed segmentation algorithm.

2 The UDWT

The DWT can be used in several slightly modified forms. The form of DWT mostly used for signal processing applications is the UDWT (Undecimated Discrete Wavelet Transform). The UDWT is also found (with very small or no difference) by the name of SWT (Stationary wavelet transform), e-decimated wavelet transform, à trous algorithm and redundant DWT etc.

The coefficients of DWT are decimated and that of UDWT are left as they are so that every level of transformation sees as many coefficients as the original signal. The drawback of a traditional DWT is that it is time-invariant, (i.e., the transform coefficients of a delayed signal are not a time-shifted version of the transform coefficients of the original signal), because of the fact that the decimation process is not time-invariant. The presence of this property makes the transform suitable for the applications in which a certain time interval in the transformed signal has to be relocated in the original signal. The convolution process preserves the time-invariance property of the discrete wavelet transform.

The implementation of UDWT analysis is carried out by omitting the decimation process while up-sampling the low and high pass filters by a factor of 2 after every level. The up-sampling of filters is needed because the DWT changes the scale of the coefficients instead of that of the filters using the decimation process, which being omitted in this case, needs to have the filters up-sampled in order to change the scale of the transform.

3 The Segmentation Algorithm

The selection of a suitable wavelet for decomposition of image led us to Coiflet orthogonal wavelet (fig. 1). This wavelet transforms the image so that its approximation subbands are in agreement to the original image to a higher degree as compared to other wavelets. As the algorithm directly maps the detected/segmented areas from the transform coefficients to the original image, this helps in getting precise results.

The next step in this process is to select a suitable wavelet level whose subbands are to be used in the proposed algorithm. Usually, this is considered to be dependant on the size of the target objects.

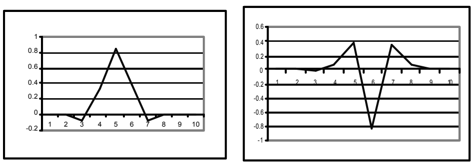

Fig.1 (a) Coiflet scaling and (b) wavelet functions

As this study responds to real-life problems, we define certain targets to be normal and others as abnormal. The normal targets are the ones that have;

1. An average grey level different from that of the background
2. An area that is defined by comparatively stronger edges
3. An internal pattern (if present) not overpowering the shape of the target

Every coefficient at level j of a wavelet transform approximation subband is some linear combination of NxN of its neighbor coefficients at level j-1; every next level shifts the area with size defined by that of the filter, towards its average grey value. This helps in segmenting such target areas that have non-uniform structure. The target objects with strong edges are mostly the areas of interest in real-life cases. The third characteristic for normal targets limits them to have a pattern that overpowers their shape i.e. the objects with very large (compared to the object's size) and strong (high contrast) pattern can be recognized as multiple objects which is misleading.

These (normal) kinds of targets were found to exist in most of the real-life problems e.g. living creatures or man-made objects in sky, fields or rough terrain. If the three conditions are met, the problem solution becomes independent of the size of the target. However as we are considering the cases that have at least one class of pixels that is a strong candidate for the background of the scene, the size of the target objects is limited to an extent up to which the background class remains the largest class in the image. The 3rd level of UDWT was thus found to be reasonable for the presented algorithm for most natural scenes and "normal" targets and is considered as the selected UDWT approximation subband.

The normalized histogram of the image can be thought of as a mixture of probabilities of all classes in the image. Therefore it is considered as its PDF (probability density function). In our case it is calculated for the selected approximation subband. The PDF calculated in this way carries singularities. The reason for this is that as the regions of interest move towards their average grayscale levels, the number of grayscale values for a region gets concentrated around a mean value (average) and the amplitude of this range gets higher and higher. This effect is observed by all regions and thus the respective PDF gets distinct peaks/singularities. This PDF is not suitable for further processing therefore it is again decomposed using UDWT into a number of levels. The previously proposed method [4] uses a high transform level (5th or higher) for detecting the greatest local minima as adaptive threshold. This method merely separates a combination of background and some target classes from other target classes. We have used separate approximation subbands of the 1-D UDWT of the PDF to estimate the background class and to find adaptive thresholds.

According to the central limit theorem, for a linear combination of N samples, as N increases, the combination approaches a normal distribution. Therefore, the 1-D UDWT of the PDF can be modeled using Gaussian mixtures as it is a linear transform. This Gaussian is a nonlinear function and thus its exact modeling is complex, however a graphical technique is developed that approximates the exact model sufficient enough for this application. Every normal distribution has a plot that has at least one side (in case of a mixture) approaching linearity for a certain length. We have manipulated this fact and experimentally calculated certain parameters that can be used to model every PDF to find its background class.

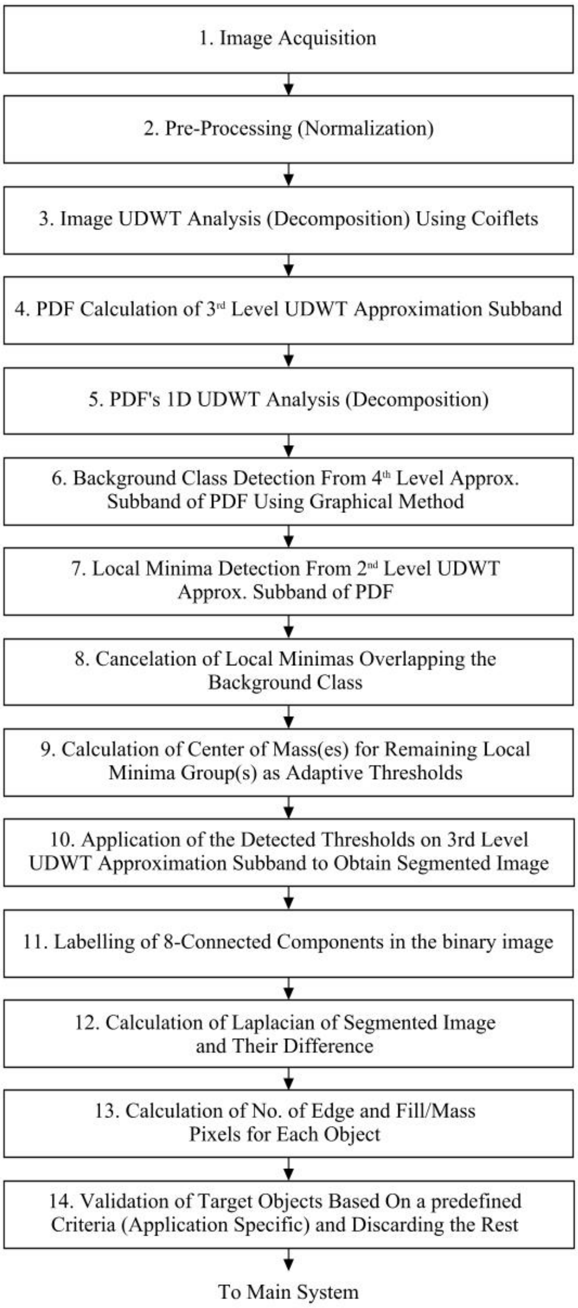

Fig. 2 - Adaptive natural scene segmentation algorithm flow

The 4th level 1-D UDWT approximation subband of the PDF is used for the following graphical method to extract the background class because up to this level all the singularities get removed. We look for segments of straight lines in the selected subband of PDF. To prove the assumption of the straight line segments for a Gaussian function, it is plotted for different values of m (mean) and s (variance) and these segments are searched.

$$f(x) = \frac{1}{\sqrt{2p}\, s_i} \exp\left[-\frac{1}{2}\left(\frac{x - m_i}{s_i}\right)^2\right] \quad (1)$$

The bell-shaped plot of this function remains approximately constant for certain duration and then the rate of change of slope starts changing. Therefore the criteria are that m and s are constant for a given distribution and $df(x)/dx$ remains approximately constant for an interval of x which is inversely proportional to $df(x)/dx$.

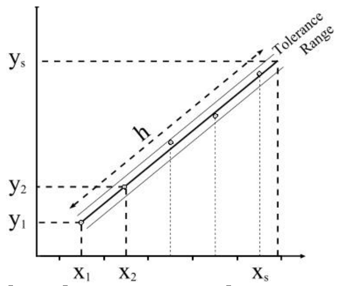

Fig. 3 - Short line segment detection from PDF using graphical method

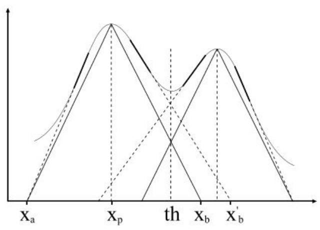

Fig. 4 - Background class estimation using triangles

The above process is similar to regression, we use every consecutive pair of $f(x)$ values to construct a straight line, then we extend this line to see where on x it stops for a preset value of h (the line segment length). Then we check all points up to the final x if they have minimum error from the constructed line. Usually this is done using the method of least squares to find the equation of line and then calculate the values of $f(x)$ for errors. But it is not applicable here since we need to find the value of x whereas the line length is kept predefined. The value of h for natural scene images is found using experimentally. In other words;

From equation of line passing through two points and figure 3;

$$\frac{b}{a} = \frac{y_s - y_1}{x_s - x_1} \quad (2)$$

Where $b = y_2 - y_1$ and $a = x_2 - x_1$, then for the right angled triangle;

$$h^2 = (y_s - y_1)^2 + (x_s - x_1)^2 \quad (3)$$

From equations (2), (3);

$$\Rightarrow x_s = x_1 + \sqrt{\frac{a^2 h^2}{a^2 + b^2}} \quad (4)$$

And from equation (2);

$$\Rightarrow y_s = y_1 + \frac{b}{a}(x_s - x_1) \quad (5)$$

Equations (4) and (5) give us the unknown point on x axis as the final limit of the line segment and y as the $f(x)$ value at that point. Since the value of x-axis index is "calculated" and we only have integer indices, this value is rounded-off. Now a tolerance range is set that accepts the $f(x)$ value inside it to be a valid approximated line which is also found experimentally.

If x_s is the very next pixel with respect to x_2, then we calculate the values of y or $f(x)$ only for x_s (as x_s is rounded-off so y_s also has to change), otherwise for all the pixel indices from x_2 up to x_s and it is checked if the actual value corresponding to these indices exist inside the tolerance limit. If the values agree to the criteria, then the points from x_1 to x_s are considered to be a line segment. This procedure is illustrated in Fig. 3.

Once the algorithm identifies a line segment (Note: The first line segment identified will have an increasing slope), it looks for a "peak", which is a point at which $f(x)$ has a slope value '0' and is greater than both of its adjacent neighbors, i.e.

$$\frac{d}{dx} f(x) = 0 \text{ and } f(x-1) < f(x) > f(x+1)$$

If these conditions are met then this point is declared as a peak and the algorithm starts looking for a straight line segment with decreasing slope. This process is the same as for the increasing slope search. The successful peaks (that are in middle of two oppositely sloped straight lines) are recorded and their corresponding increasing and decreasing slope line segments are used to formulate the parameters of a triangle. We replace the line with the smaller slope with sign inverted greater slope (the reason for this is that a smaller slope only occurs in case this Gaussian distribution is mixed with another one and we want to calculate a single distribution). Smaller slope rejection;

$$\text{if } x_b < x_a \text{ then } x_b = (x_p - x_a) + x_p = 2x_p - x_a \quad (6)$$

$$\text{if } x_a < x_b \text{ then } x_a = x_p - (x_b - x_p) = 2x_p - x_b \quad (7)$$

The triangle has the x-axis as base, limited by its intersections with the extensions of the line segments, and the two extended lines as the other two sides. Then the area of triangle can be calculated. We only record the x_1 value if a line segment is detected therefore at this time we calculate their intersections with x-axis (x_a for increasing slope, x_b for decreasing). From figure 4;

$$x_a = x_{1i} - f(x_{1i}) \left[\frac{(x_{1i} + 1) - x_{1i}}{f(x_{1i} + 1) - f(x_{1i})} \right] \quad (8)$$

Similarly;

$$x_b = x_{1d} - f(x_{1d}) \left[\frac{(x_{1d} + 1) - x_{1d}}{f(x_{1d} + 1) - f(x_{1d})} \right] \quad (9)$$

Where x_{1i} is the starting point for the increasing slope line and x_{1d} for that of decreasing slope, Area of triangle;

$$\Delta = (x_b - x_a) \left[\frac{f(x_p)}{2} \right] \quad (10)$$

Then the areas of all detected triangles are compared to find the largest. The x-axis limits of this triangle define the range for the background class.

The 2nd level approximation subband of the PDF 1-D UDWT is then used for local minima detection. The coefficients of this level have comparatively sharper structure as compared to 2nd level which was used in previous procedure. Therefore there exist many local minima in most cases. The local minimum is defined by the inequality;

$$f(x-1) > f(x) < f(x+1) \quad (11)$$

The magnitude of each local minimum is also recorded with its location which is used in the following step. The local minima that are inside the range of that of the detected background class are then cleared. This is done due to the fact that no threshold can exist inside this range.

If the background class has limits in the PDF such that they are bounded by groups of local minima (after cancellation) then two thresholds are calculated. There can also be cases where only one group exists. The proposed adaptive threshold calculation is done using the center of mass equation from the detected local minima;

$$R_{CM} = \left[\sum_{i=i_{\min}}^{i_{\max}} [i.f(i)\hat{u}_x] \right] \Big/ \left[\sum_{i=i_{\min}}^{i_{\max}} f(i) \right] \quad (12)$$

The amplitude is analogous to mass in this case. The thresholds found using the above method have shown extraordinary segmentation results. These thresholds are then applied to the 3rd level image UDWT using the Bayes Classifier. If $A^{(3)}(m,n)$ is the 3rd level image approximation subband, l_1 and l_2 are the center of masses of lower and higher minima group;

$$I_{seg}(m,n) = \begin{cases} 1 & A^{(3)}(m,n) < l_1 \\ 1 & A^{(3)}(m,n) > l_2 \\ 0 & l_1 < A^{(3)}(m,n) < l_2 \end{cases} \quad (13)$$

The final I_{seg} is a binary image and is then passed through 8-connected component labeling algorithm so that all the segments are labeled and total segment count t is acquired.

The validation step is the final step to accept wanted segments and discard unwanted segments. The proposed validation scheme calculates the Laplacian of the segmented image to get an edge map;

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \quad (14)$$

and then its difference to the segmented image so that a object mass map of the segmented image is acquired;

$$I_{mass}(m,n) = I_{edge}(m,n) - I(m,n)$$

The effective areas of edge and mass for each object t are then calculated as;

$$A_{edge} = \sum_{i=1}^{m*n} I_{edge}(i) \quad , \quad A_{mass} = \sum_{i=1}^{m*n} I_{mass}(i) \quad (15)$$

These maps are then multiplied one by one to the $D^{(3)}(m,n)$ (3rd level UDWT detail subband of image) to get the detail factors for edge and mass of the segmented image as I'_{edge} and I'_{mass} respectively for each object. The overall detail for each object is then calculated using these detail factor images as;

$$A'_{edge} = \sum_{i=1}^{m*n} I'_{edge}(i) \quad , \quad A'_{mass} = \sum_{i=1}^{m*n} I'_{mass}(i) \quad (16)$$

Finally the ratios for each object;

$$R_{edge} = \frac{A'_{edge}}{A_{edge}} \quad , \quad R_{mass} = \frac{A'_{mass}}{A_{mass}} \quad (17)$$

These ratios for every object are the relative "detail" content for each objects edge and mass and have been found exceptionally efficient to validate or reject an object based on its predefined/expected detail information (i.e. application specific validation).

4. Conclusion/Case Studies

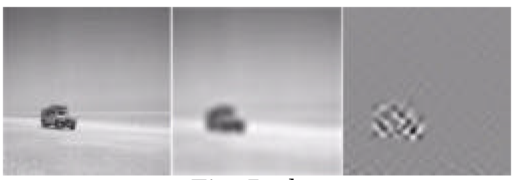

Fig. 5.a.b.c

Fig. 6.a.b.c

Object No.	*Centroid Object*		*Pixel Count*	*Type*
	X Axis Position	*Y Axis Position*		
1	9.1115	57.113	1982	Sky
2	79.628	36.416	341	**Target**
3	82.875	19.125	8	Water
4	87.787	85.787	324	Water

Table 1 – Results after connected component labeling

No.	*Edge Area*	*Edge Detail*	*Mass Area*	*Mass Detail*	*Edge Ratio*	*Mass Ratio*
1	147	154.91	1835	1757.8	1.0538	0.95793
2	90	1998.8	251	7238.4	22.209	28.838
3	8	110.23	0	0	13.778	0
4	150	858.14	174	402.07	5.7209	2.3108

Table2 – Results after ratio calculations

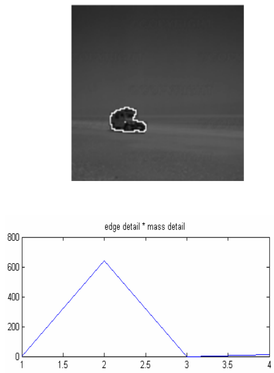

Fig. 7.a.b

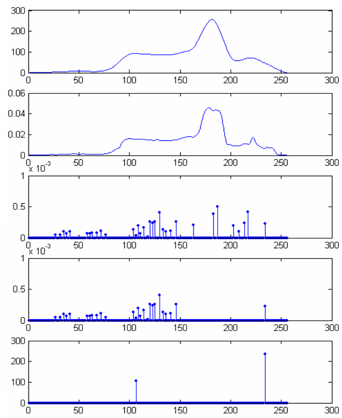

Fig. 8.a.b.c.d.e

Detected adaptive thresholds: $l_1 = 107$, $l_2 = 230$

Fig. 5 shows original image with its 3rd level UDWT approximation and detail subbands. Fig. 6 shows thresholded image based on the thresholds found from PDF 2nd level approximation (fig.8.e). Fig. 8.a is that of 4th level which is used to find the background class and discard the overlapping local minima (fig. 8.d). Fig.8.e shows the center of masses of groups of fig. 8.d. Figure 6.c shows the labeled image and the 2nd object is identified as target. Graph of fig. 7.b shows 2nd (target) object clearly discriminated. Thus 7.a shows accurate segmentation results from the presented method.

5. REFERENCES

[1] R. M. Rao and A. S. Bopardikar "Wavelet Transforms", Pearson Education, 2004

[2] T. Kubota, C. O. Alford and M. B. Woods, "Design and Implementation of Orientational Filters Using Separable Approximation Methods", Proceedings of IEEE/SMC International Conference, vol. 2, pp. 333–8, Oct. 1993

[3] Rafael C. Gonzalez and Richard E. Woods, "Digital Image Processing", Pearson Education, 2004

[4] Xiao-Ping Zhang and Mita D. Desai, "Segmentation of Bright Targets Using Wavelets and Adaptive Thresholding", IEEE Transactions on Image Processing, vol. 10, no. 7, July 2001

[5] J. S. Weszka and A. Rosenfeld, "Threshold evaluation techniques," IEEE Trans. Syst., Man, Cybern., vol. SMC-8, no. 8, pp. 622–629, 1978.

[6] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," IEEE Trans. Pattern Anal. Machine Intell., vol. 11, no. 7, pp. 674–693, July 1989.