UDC:

Automatic and Robust Object Recognition in a Video Minimizing the Gradient Energy function

N. Krishnan, P. Vijaya lakshmi, S. Naga Nandhini Sujatha

1. Senior Member, IEEE, Manomanium Sundaranar University, Tirunelveli Tamil nadu, India
2 Members IEEE, Kamaraj College of Engg and Technology , Virdhunagar. Tamil Nadu, India
3. Member IEEE, K.L.N. College of Engg, Madurai. Tamil Nadu, India .
krishnan@computer.org , vijikcet@gmail.com, nandhini_sujatha@rediffmail.com

Abstract.

The automatic image analysis and understanding of visual content is a challenging problem in computer vision. Despite we have a number of detection processes, ranging from template matching to flexible methods; there is an urge for a single robust algorithm to detect any arbitrary object irrespective of geometric changes, automatically. Images from real time applications are corrupted by noise from various resources. Here, the bilateral filter smoothes the images while preserving edge, in a way that is tuned to human perception. This paper presents a supervised learning technique of automatic recognition of arbitrary object in a multi object class environment, employs the extended gradient vector field (EGVF) which demands no human intervention for initial contour condition and has high resistance to noise; then, a robust Iris adopter operates on these gradient fields for classification of pixels . To locate and detect any arbitrary object unique patterns of the query object using adjacency matrix is employed. Lower probabilities of error classification are obtained by using contour sequence moments than area based moments. Major Applications include target detection from radar images and automation using robot vision.

Keywords: *Bilateral filter, EGVF, Iris Filter, Detection, Patterns, Contour Moments invariants, Supervised Learning.*

1. Introduction

Image segmentation and object detection plays an important role in applications where image analysis or understanding is necessary. Traditional snake algorithms have a drawback that the construction of initial contour often requires human interaction. Yuen et al. [1 proposed an automatic initialization snake algorithm for multi object segmentation. Their algorithm is limited to some specifically distributed objects, e.g., those spreading around the center of gravity. Bilateral filter was introduced by Tomasi and R.Manduchi which will give edge-preserved smoothening process [2].Bilateral filtering is a nonlinear filter that considers intensity variations as well as spatial closeness in the noise smoothing process [3]. For high level segmentation, that means extracting objects from back ground, we need good detection of object boundaries. The active contour model [4], also called as traditional snake algorithms provides good results but the initial contour often requires human interaction. The popular watershed algorithm is a powerful tool for image segmentation but it suffers with over segmentation problem. Differential method of edge detection uses approximation of spatial gradient at each pixel location, where as in gradient method, four directional gradient vector flow (GVF) [5] for each pixel in the image is calculated. To obtain eight directional gradient vector flow of each pixel, Extended GVF [6] is employed. The Iris Filter is used for the detection of approximately rounded convex regions. The filter performance does not depend on the contrast; thus detection in very weak contrast images can be made clearly. Then region-filling technique is used to separate the foreground objects detail. [7].

Foreground object color cluster extraction at the initial frame using K-means algorithm and filtering via Foreground Extraction Mask is discussed [8]. A rudimentary detector is learned from a very small set of segmented images and applied to a larger training set of un segmented images. The detectors are learned with a boosting algorithm which creates a location-sensitive classifier using a discriminative set of features from a randomly chosen dictionary of contour fragments[9] An adaptive foreground object extraction algorithm for real-time video surveillance is presented. The proposed algorithm improves the previous Gaussian mixture background models (G M Ms) by applying a two-stage foreground/background classification procedure to remove the undesirable subtraction results due to shadow, automatic white balance, and sudden illumination change [10]. Under the proposed framework, a novel algorithm for detecting foreground objects from complex environments is then established [11]. A new method is presented to learn object categories from unlabeled and un segmented images for generic object recognition. We assume that each object can be characterized by a set of typical regions, and use a new segmentation method - "Similarity-Measure Segmentation" - to split the images into regions of interest [12]. Another paper presents an efficient technique for extracting closed contours from range images edge points. The proposed approach consists of three steps. Initially, a partially connected graph is generated from those input points. Then, the minimum spanning tree of that graph is computed. Finally, a post processing technique generates a single path through the regions' boundaries by removing noisy links and closing open contours [13]. Artificial Neural Network (ANN) based scheme for object recognition is presented. The trained neural network is used to predict the actual moments from the moments of the object with different occlusions. The object is recognized based on the comparison of the actual moments and the predicted moments [14].

2. Proposed Work:

In computer vision applications, segmentation and object detection are the key steps. The proposed work is focused to identify objects by extracting the foreground details from the image by computing the extended gradient vector fields (EGVF) for every pixel and minimizing the gradient energy function. The Extended GVF preserves the gradient near object boundaries and diffuses them in the areas of low intensity. The EGVF field is adopted as the energy term which is more robust to noise than the traditional image gradients. Generally, Iris filters are used to extract approximately convex regions in the images. Here, we apply the Iris filter to classify the pixels as "object and non object" by operating over the gradient vectors fields. Since this work is featured for automatic detection in a multi class objects and for real environment, patterns are used for generic object recognition. The fist of this method lies on the representation that has a sufficient descriptive power to allow discrimination between dissimilar objects. To locate and detect any query object in the foreground, unique patterns of the query object generated using adjacency matrix, is employed. Here, we compute translation, rotation and scale invariant normalized contour sequence moments for shape description of objects. Recognition is implemented by classifying the feature vectors using Euclidean distance matching of mean proto type feature vector of known classes and unknown object class by statistical classifier under the supervised learning approach.

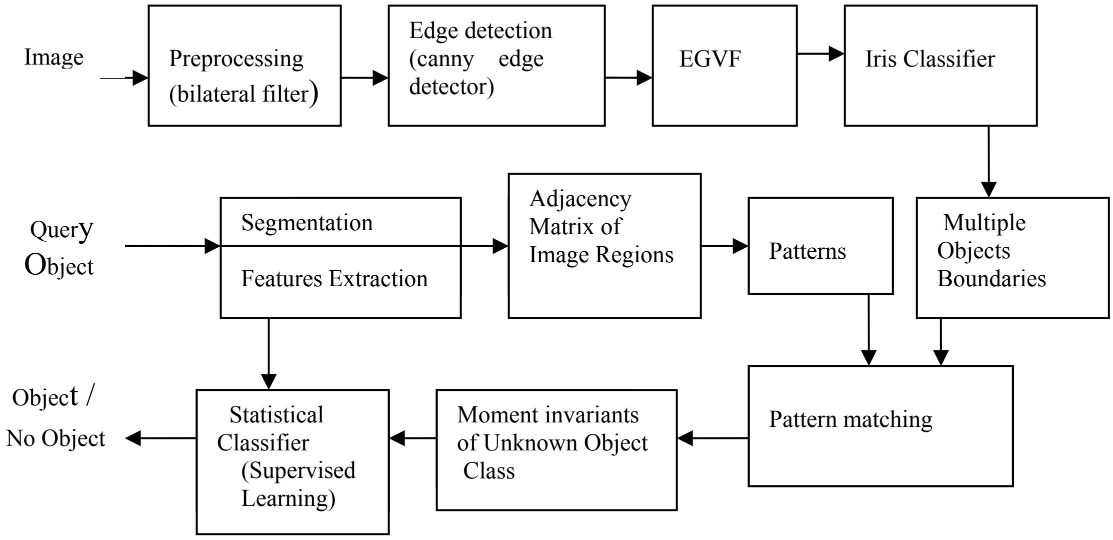


Figure 2.1 Architecture of Proposed Work

3. Foreground details extraction.

3.1. Preprocessing using bilateral filter:

Linear filtering is implemented by convolution while nonlinear filtering does not demand any predefined function for smoothening process. Here, preprocessing deals with techniques for enhancing contrast, removing noise and isolating regions of texture. Edge detection is achieved by canny edge detector; since its edge maps are much useful to achieve more robust object tracking. In order to reduce the texture effect and non-object lines, the image is preprocessed by employing the bilateral filter [15], [16]. Bilateral filtering smoothes images while preserving edge, in a way that is tuned to human perception by means of a nonlinear combination of nearby image values. It combines gray levels or colors based on both their geometric closeness and their photometric similarity, and prefers near values to distant values in both domain and range. Texture removal becomes an important preprocessing [17] when objects in a scene have a textured back ground, since texture often contains a high density of edges.

3.2. Extended Gradient Vector Flow Model:

GVF field is the external gradient force to solve the problem of sensitivity to initial contour condition. The GVF field is defined as the force field of vectors. For every image pixel f(x, y), $V(x,y)=[u(x,y),v(x,y)]$ is computed by minimizing the energy function as,

$$\varepsilon=\iint \mu(u_x^2+u_y^2+v_x^2+v_y^2)+|\nabla f|^2|V-\nabla f|^2\,dxdy$$

where ∇f is the gradient of the edge image derived from the original intensity image f, μ is a regularization parameter, and the subscripts represent partial derivatives with respect to x and y axis. The traditional GVF field is stable only with the Courant-Friedricchs-Lewy step-size condition [6], the regularization parameter μ should be restricted to 0 ~ 0.25. After the minimization process, V (x, y) will approximate∇f, and it is large near object boundaries and smoothed out elsewhere. Each GVF vector will point toward object boundaries even if they are from them. The GVF field is adopted as the external energy term for its stronger resistance to noise than the traditional image gradients. In our method, a force field of four components EGVF vectors are defined, which represent the flow amplitudes and directions respectively shown in the Figure 3 .1. Here, (x, y) and (x1, t1) constitute two sets of orthogonal coordinate systems with a rotating angle of 45^0.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

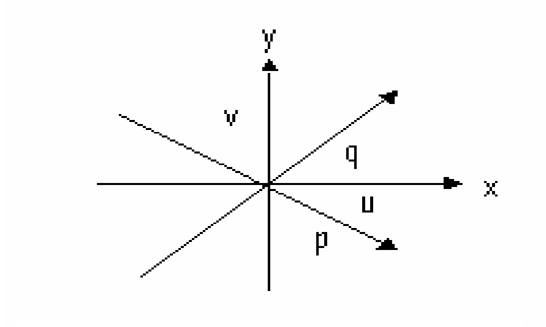


Figure 2.2 Definition of force Directions

As the extension of original GVF field mode [6], this force vector field can be described as

EGVF(x, y) = $[V_1(x, y), V_2(x', y')]^t = [(u, v), (p, q)]^t$ which minimizes the energy functional,

$$E = \iint \mu(|\nabla u|^2 + |\nabla v|^2) + |\nabla f|^2 |V_1 - \nabla f|^2 dxdy + \iint \mu(|\nabla' p|^2 + |\nabla' q|^2) + |\nabla' f|^2 |V_2 - \nabla' f|^2 dx' dy'$$

Here the gradient operators are defined as Δ= (∂/∂x, ∂/∂y) and Δ'= (∂/∂x', ∂/∂y')

3.3 Iris Filter as a Robust Classifier:

Object and non-object pixels can be isolated as local low density areas on the image. But their absolute values are not constant and vary in size due to background variations, imaging conditions and so on. So, we employ an iris Filter for classification purpose, whose performance do not depend on the contrast, thus objects in images of very weak contrast can be detected clearly. The flow amplitude and directions of gradient vectors of the edge pixels are used to compute the kernel of the iris filter. The kernel operates over the gradient energy field to classify the pixels. The sum of cosine of differences of angles equal to 1 then the pixel is an object pixel, otherwise it is a non object pixel.

3.4 Algorithm for Automatic foreground Details Extraction:

1. Read the image $f(x, y)$
2. Pre process the image using bilateral filter to remove the noise and to isolate regions of texture
3. Obtain the Edge maps $f_{e(x,y)}$ of the enhanced image using canny edge detector.
4. Calculate the Extended gradient vector field for each pixel of the edge image $f_{e(x,y)}$,

   $V(x, y) = [u(x, y), v(x, y), p(x, y), q(x, y)]$
5. Minimize the energy function, E, so that ∇f is large near objects boundary and

minimum else where.

$$E = \iint \mu(|\nabla u|^2 + |\nabla v|^2) + |\nabla f|^2 |V_1 - \nabla f|^2 \, dxdy + \iint \mu(|\nabla' p|^2 + |\nabla' q|^2) + |\nabla' f|^2 |V_2 - \nabla' f|^2 \, dx'dy'$$

6. Calculate the Iris filter Kernel using gradient vector flow magnitude and directions
7. Apply the Iris filter over the gradient vectors to classify the pixels as object and non object pixels, $f_{if}(x, y)$

4. Object detection and recognition

4.1 Pattern Generation:

To detect the query object, first unique patterns of the object are generated as shown in the Figure. Then a directed search over the video is done to locate the object. In real images, patterns are corrupted by noise, geometric distortion occlusion etc. So, an absolute equivalent is not possible. Patterns are of small, they represent whole objects of interest. To resolve the variable object size, let the query image has the dimensions $m_q \times n_q$ and that this represents the smallest possible size of the object in a search image. The largest scaling of the query image 'g' that fits into the search image 'f', while preserving the aspect ratio, is determined,

$$\gamma_H = \min\left(\frac{m_f}{m_q}, \frac{n_f}{n_q}\right) \tag{4.1}$$

where $m_f \times n_f$, are dimensions of search image. We search for objects of arbitrary size by considering multiple scale factors between 1.0 and γ_H.

Figure 3.1 Pattern Generation

4.2 Contour Sequence Moment Invariants:

Moments invariants are useful for shape analysis. They can also be computed optically at high speeds. Moment invariants have been used in distinguishing between shapes of different aircraft, character recognition, and scene matching applications. A closed boundary is characterized by an ordered sequence $z(i)$ that represents the Euclidean distance between the Centroid and all N bounding pixels of the digitized shape. No extra processing is required for shapes having spiral or concave contours. The r^{th} contour sequence moment m_r and the r^{th} central moment μ_r are calculated,

$$m_r = \frac{1}{N} \Sigma \sum_{i=1}^{N} [z(i)]^r \tag{4.2}$$

$$\mu_r = \frac{1}{N}\sum_{i=1}^{N}[z(i) - m_1]^r \quad (4.3)$$

Translation, rotation and scale invariant normalized contour sequence moments $\overline{m_r}$ and $\overline{\mu_r}$ are calculated [18] and these are used directly for shape representation.

$$\overline{m_r} = \frac{m_r}{(\mu_2)^{\frac{r}{2}}} = \frac{\frac{1}{N}\sum_{i=1}^{N}[z(i)]^r}{[\frac{1}{N}\sum_{i=1}^{N}[z(i) - m_1]^2]^{\frac{r}{2}}} \quad (4.4)$$

$$\overline{\mu_r} = \frac{\mu_r}{(\mu_2)^{\frac{r}{2}}} = \frac{\frac{1}{N}\sum_{i=1}^{N}[z(i) - m_1]^r}{[\frac{1}{N}\sum_{i=1}^{N}[z(i) - m_1]^2]^{\frac{r}{2}}} \quad (4.5)$$

Lower probabilities of error classification are obtained by using contour sequence moments than area based moments. Less noise sensitive shape descriptors are calculated,

$$F_1 = \frac{(\mu_2)^{\frac{1}{2}}}{m_1} \quad (4.6)$$

$$F_2 = \frac{\mu_3}{(\mu_2)^{\frac{3}{2}}} \quad (4.7)$$

$$F_3 = \frac{\mu_4}{(\mu_2)^2} \quad (4.8)$$

Object Recognition is implemented based on the use of decision (or discrimination) functions. Let $X = (x_{1,} x.......x_n)^T$ represent 'n' dimensional pattern vector. For W pattern classes $w_{1,} w_{2,} w_3........w_w$ decision theoretic pattern recognition is to find W decision functions $d_1(x), d_2(x),........d_w(x)$ with the property that if a pattern x belongs to class $w_{i,}$ then,

$$d_i(x) > d_j(x) \text{ , j = 1,2,…W; j \#i} \quad (4.9)$$

Recognition based on matching represent each class by a pro type pattern vector, is shown in the Figure 3.2. An unknown pattern is assigned to the class to which it is closest in terms of distance. We used minimum distance classifier; it computes the Euclidean distance between and each of the pro type vectors. It chooses the smallest distance to make a decision. We define the pro type of each pattern class to be the mean vector of the pattern class,

$$m_j = \frac{1}{N_j}\sum_{x\in\omega_j} X_j \text{ , } \quad j = 1, 2,\dots W \quad (4.10)$$

here Nj is the number of pattern vector from class w_j and the summation is taken over these vectors. W is the number of pattern classes. One way to determine the class membership of an unknown object 'x' is to assign it to the class of its closest pro type. The Euclidean distance is used to determine the closeness. Evaluating the smallest distance is equivalent to evaluating the function.

$$d_j(X) = X^T m_j - \frac{1}{2} m_j{}^T m_j \text{, } j = 1, 2,\dots W \quad (4.11)$$

and assigning X to class ω_i, if $d_j(X)$ yields the largest numerical value. The decision boundary between classes ω_i and ω_j for a minimum distance classifier is

$$d_{ij}(X) = d_i(X) - d_j(X) = X^T(m_{i\_}m_j) - \frac{1}{2}(m_i - m_j)^T(m_i - m_j) \tag{4.12}$$

4.3. Algorithm for Object Recognition:

1. *$f_{if}(x, y)$ be the Fore ground details of the video image*
2. *g be the query object image*
3. *Extract the transform invariant features of the query object image.*
4. *Generate the unique pattern of the query object image*
5. *Evaluate matching criteria to locate the query object in the fore ground video image $f_{if}(x, y)$*
6. *Extract the unknown object window from the video image $f_{if}(x, y)$ based on the location of the local maxima Criteria.*
7. *Derive the transform invariant features of the unknown object window.*
8. *Calculate the matching criteria for the features of query object image and unknown object window using supervised learning method.*

5. Analysis of Experimental Results:

RSI's IDL has been applied for implementation. It provides composite data types such as character strings, homogeneous-type arrays, and non-hierarchical record structures of mixed data types. As most other array programming languages, IDL is very fast doing vector operations. The European Space Agency used RSI's IDL to process almost all of the pictures of Halley's Comet taken by the Giotto spacecraft. Here,all he experiments are carried out with twenty model objects from a video to demonstrate its efficiency of recognition using the RSI' s IDL version 6.3 Win 32 (Interactive Data Language) as it was taking pretty very less time for recognition. The computational efficiency of the program is tested for various image types are shown in the Table 5.3. Also, the efficiency of the algorithm is tested with various kernel size of Iris Filter and the results are shown in the Table 5.4. The 60 test cases were generated from a video database of 100 model objects by translating, scaling and rotating the model objects. The test objects were randomly rotated and translated, but scaled to factor of around $\frac{3}{4}^{th}, \frac{1}{2}$, 2, 3 and some without scale of their model sizes. In case of testing occlusion, the model objects are translated, scaled, rotated and occluded. The occluded objects are generated by occluding the model objects with little or no change in center of gravity. The results of the proposed work are compared with the classical Moments based detection is shown in the Table 5.2. In case of testing the noisy objects, the noise is added using the impulse noise with a given percentage density. The result of the new technique are compared with the classical moments based detection shown in the Table 5.1

Percent Noise (%)	Recognition rate in (%)	
	Classical Work (Moment Invariants)	Proposed Work
10	59	86
20	48	72
30	41	70
40	15	64

Table 5.1 Effect of Noise on Recognition Rate

Methods	Similarity Transformations (60 Test Cases)	Noise + Similarity (60 Test Cases)	Occlusion (60 Test Cases)	Noise + Occlusion (60 Test Cases)
Classical Work (Moment Invariants)	60	40	0	0
Proposed Work	60	60	57	52

Table 5.2 Similarity for 60 Test Cases

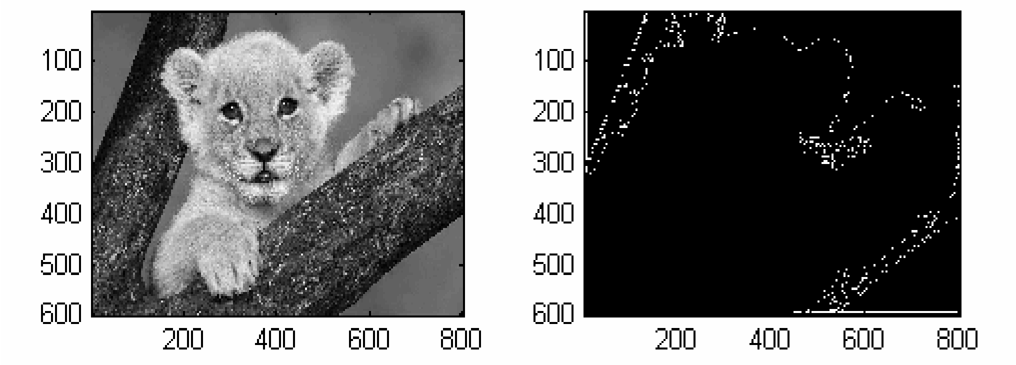
Coins. png		Parrot. jpg	
Kernel Size	Time in seconds	Kernel Size	Time in seconds
5x5	5.9	5x5	9
7x7	7	7x7	12
11x11	13	11x11	15.4

Table 5.3 CPU Efficiency for Various Images

Image Type	Image Size(bits)	Time taken in (Sec)
Parrot. jpg	600x800	12
Cubbie. jpg	600x800	11
Coins. png	600x800	7
Rice. png	600x800	7.2
	Average	9.3

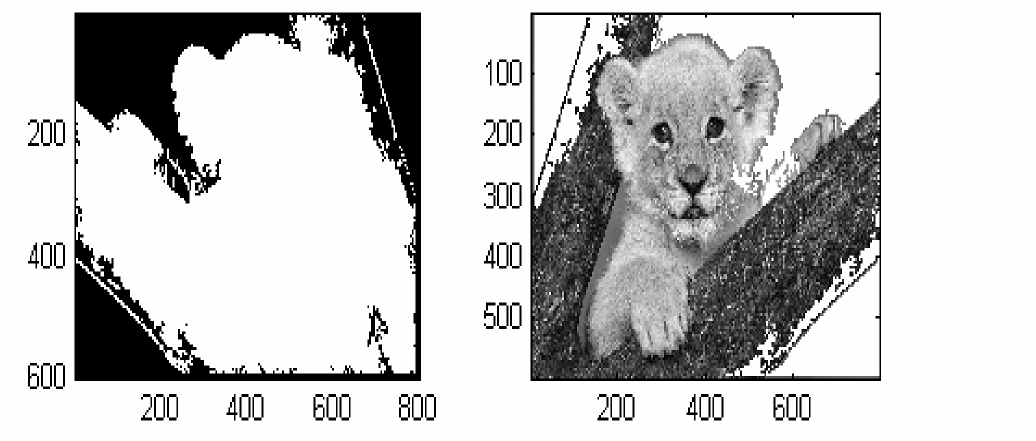
Table 5.4 CPU Efficiency with Iris Filter Kernel Size

The experiments are also conducted with natural and real images. The results of the experiments are shown in the Figure 6.1, 6.2, 6.3 and 6.4.



(a) **(b)**

Figure 5.1 a) original Image b) Edge image



a) **b)**

Figure 5.2. a) Iris Filtered Image b) Detected

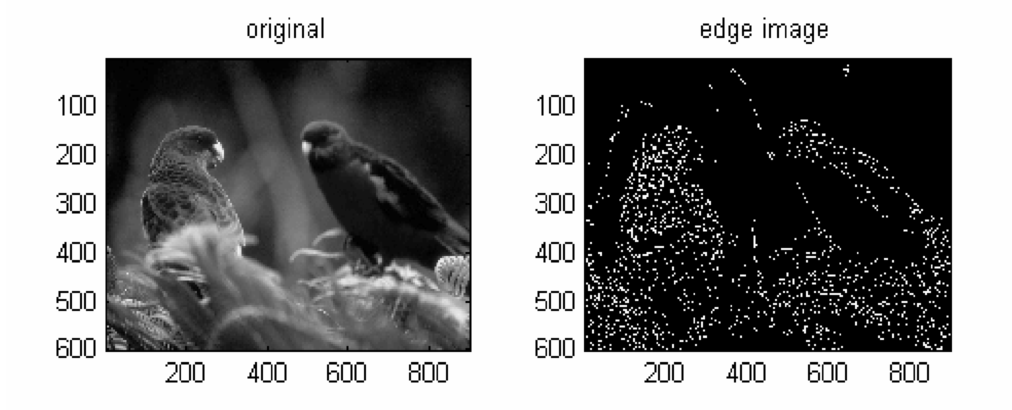


Figure 5.3. a) original Image b) Edge Image

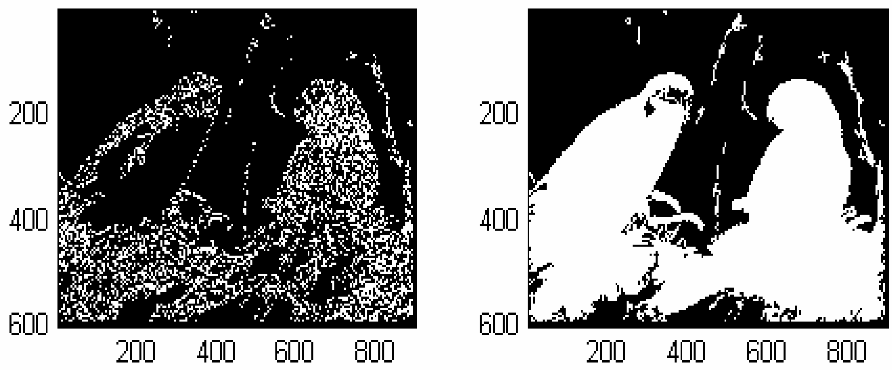


Figure 5.4 a) Iris Filtered Image b) Detected

6. Conclusion:

This paper presents a fast approach for automatic detection of objects in a video which is highly insensitive to noise and also relives the burden of initial contour condition to detect the objects boundaries. The Iris adopter whose performance is independent of contrast makes the system more robust. Detection of objects by unique patterns using adjacency matrix and normalized contour sequence moment invariants facilitate the work to identify any arbitrary object irrespective of geometric changes. This work can be augmented by training the artificial neural networks under unsupervised learning approach.

7. References:

[1] X.Liu and D.L.Wang. "A spectral histogram model for texton modeling and texture discrimination," Vis, Res, Vol.42, no.23, pp 2617-2634-2002. 2005

[2] C. Tomasi and R. Manduchi Proceedings of the 1998 IEEE International Conference on Computer Vision.

[3] "Multimedia object retrieval for image database "IEEE Trans on Image Processing Dec 2006.

[4] W.N.Lie and C.H.Chuang, "A fast and accurate snake model for object contour detection," Election, Lett., Vol. 37. no.10, pp .624-626, 2001.

[5] Chenyang Xu and Jerry L. Prince, "Snakes, Shapes, and Gradient Vector Flow", IEEE Transactions on image processing, vol. 7, no. 3, march 1998

[6] Cheng-Hung Chuang and Wen-Nung Lie, "A Downstream Algorithm based on Extended Gradient Vector Flow field for object segmentation" IEEE Transactions on image processing, vol. 13, no. 10, Oct 2004.

[7] Kobatake, H.; Murakami, M. Adaptive filter to detect rounded convex regions: irisfilter Pattern Recognition, 1996. Proceedings of the 13th International Conference on Volume 2, Issue , 25-29 Aug 1996 Page(s):340 - 344 vol.2

[8] Richard Y. D. Xu, John G. Allen, Jesse S. Jin Robust Real-Time Tracking of Non-rigid Objects School of Information Technologies University of Sydney, NSW 2006, Australia {richardx,jallen,jesse}@it.usyd.edu.au

[9] Jamie Shotton ,Andrew Blake, Roberto Cipolla Contour-Based Learning for Object Detection,Department of Engineering University of Cambridge jdjs2@cam.ac.uk, Microsoft Research Ltd. Cambridge, UK, blake@microsoft.com Roberto Cipolla Department of Engineering University of Cambridge cipolla@eng.cam.ac.uk, 2005

[10]. A Hui-Chi Zeng; Shang-Hong Lai , Adaptive Foreground Object Extraction for Real-TimeVideo Surveillance with Lighting Variations, Acoustics, Speechand Signal Processing, 2007. ICASSP2007. IEEE International Conference. Volume1, Issue , 15-20 April, 2007. Page(s):I-1201-I-1204 Digital Object Identifier 10.1109/ICASSP.2007.366129.

[11] Liyuan Li; Weimin Huang; Irene Yu-Hua Gu; Qi Tian Statistical modeling of complex backgrounds for fore ground object detection, Image Processing, IEEE Transactions Volume 13, Issue 11, Nov. 2004 Page(s):1459-1472, Digital Object Identifier 10.1109/TIP.2004.836169

[12] Michael Fussenegger1,2, Andreas Opelt1,2, Axel Pinz2 and Peter Auer1. Object Recognition Using Segmentation for Feature Detection 1. Institute of Computer Science, University of Leoben, Austria {auer, andreas.opelt}@unileoben.ac.at 2 Institute of Electrical Measurement and Measurement Signal Processing, TU Graz, Austria {fussenegger, opelt, pinz}@emt.tugraz. 2006.

[13] Angel Domingo Sappa, *Member, IEEE* Unsupervised Contour Closure Algorithm for Range Image Edge-Based Segmentation IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 15, NO. 2, FEBRUARY 2006.

[14] Saibal Dutta and P.K.Nanda Moment based Object Recognition using Artificial Neural Network "National Conference on Intelligent Systems" Image processing and computer vision Lab, Department of Electrical Engineering, National Institute of Technology, Rourkela-769008, Orissa, India Aug 2007 saibal.13@gmail.com, pknanda d13@rediffmail.com

[15] T.Randen and J.H.Husoy, "Filtering for texture classification: A comparative study," IEEE Trans. Pattern Anal.Mach.Intell., Vol.21, no.4, pp.291-310, Apr.1999

[16] D.A.Clausi and H.Dend, "Design- based texture feature fusion using Gabor filters and co-occurrence probabilities." IEEE Trans, Image Process, vol.14, no.7,pp, 925-936, jul.2005

[17] Xiuwen Liu and DeLiang Wang, "Image and Texture Segmentation using Local Spectral Histograms", Transactions on image processing, vol. 15, no. 10, Oct 2006.

[18] Gupta and Srinath Contour based moments for classification of closed planar shapes, Pattern recognition, 20(3): 267-272, 1987.

Number of Figures/ Diagrams: 7
Number of Tables: 4

Article received: 2008-01-06