MULTIPLE OBJECT DETECTION AND TRACKING IN SONAR MOVIES USING AN IMPROVED TEMPORAL DIFFERENCING APPROACH AND TEXTURE ANALYSIS

Tudor BARBU[1]

Lucrarea de față abordează domeniul detectării şi urmăririi obiectelor video în secvențele de tip sonar. Într-o primă fază sunt aplicate anumite operațiuni de preprocesare a secvenței video analizate, precum o eliminare a zgomotului bazată pe ecuații cu derivate parțiale şi o îmbunătățire a contrastului. În continuare regiunile supuse mişcării sunt detectate automat prin intermediul unei metode îmbunătățite de diferențiere a cadrelor. Câteva operații matematice morfologice sunt utilizate în scopul identificării principalelor obiecte aflate în mişcare. O tehnică de urmărire a obiectelor video multiple, utilizând o analiză a texturii bazată pe filtrarea Gabor 2D a obiectelor imagistice şi un proces de potrivire a acestora, este apoi propusă.

This paper approaches the computer vision task of video object detection and tracking in sonar movie sequences. Some pre-processing operations, such as a PDE based video denoising and a contrast enhancement, are performed on the analyzed sequence, first. Then, the moving regions are detected automatically using an improved frame-difference based approach. Some mathematical morphological operations are used to identify the main moving objects. A multiple video object tracking technique, using a 2D Gabor filter-based texture analysis of image objects and an object matching process is then provided.

Keywords: sonar video sequence, PDE image denoising, moving object, object detection, temporal differencing, morphological operations, 2D Gabor filtering, texture analysis, template matching.

1. Introduction

Video object detection and tracking represents an important and challenging computer vision domain. Obviously, it consists of two closely related video analysis processes.

Thus, video object detection involves locating an image object in the frames of a video sequence, while video tracking represents the process of monitoring the video object spatial and temporal changes during the movie sequence, including its presence, position, size, shape, etc. A video object tracking

[1] Senior Researcher II, Institute of Computer Science, Romanian Academy, Iasi branch, Romania, e-mail: tudbar@iit.tuiasi.ro

approach has to solve the temporal correspondence problem that is the task of matching the target object in successive video frames [1].

Usually, the tracking process starts with detecting the initial instance of the video object, then identifying that image object repeatedly in subsequent frame sequence. Video tracking is often a difficult process, due to some factors such as abrupt object motion, object occlusions and camera motion.

Numerous video detection and tracking techniques have been developed in recent years. Object detection can be performed using many approaches, such as: region-based image segmentation [2,3], background subtraction [4], temporal differencing [5], active contour models [6], generalized Hough transforms [7] and template matching [8]. Also, video object tracking has been approached through various techniques, such as: statistical methods, like those based on Kalman filtering or Hidden Markov Models [9], correlation-based object matching, kernel tracking [10], optical flow [11] or contour tracking.

Object identification and tracking has a wide variety of computer vision application areas. The most important of them are the video compression, video surveillance, human-computer interaction, video indexing and retrieval, medical imaging, augmented reality and robotics.

We propose an automatic video object detection and tracking system for multiple moving objects in sonar video sequences in this article [12]. The automatic character of the system means that video objects are detected and tracked automatically, no interactivity being present. For example, the user will not be asked to select an object from a frame to be tracked by the system. Also, the proposed system works for moving objects only and not any video objects. Thus, our video analysis techniques are applied on fixed camera movie sequences.

Also, the proposed approaches provide the best detection and tracking results when applied on ultrasound videos [12]. The sonar movie sequences considered for detection and tracking represent video scenes containing several objects (more than one) moving on a quite empty background.

The ultrasound video sequence must pre-processed before video analysis, using some noise removal and frame enhancement operations which are described in the next section [13,14]. Then, a novel moving object detection approach is proposed in the third section. It uses an improved temporal differencing technique that represents the main contribution of this paper.

The video tracking approach provided in the fourth section is based on a novel content-based object matching technique. The Gabor filtering based texture analysis approach used by the object matching process is another original and important contribution of this article [15]. The performed experiments are mentioned in the fifth section and the paper ends with a section of conclusions.

2. Sonar movie preprocessing

We proposed some sonar image and video denoising techniques in our previous works [13,14]. Thus, we developed a PDE – based speckle noise removal model that is expressed as follows:

$$\begin{cases} \dfrac{\partial I_i}{\partial t} = -\Delta\left[f\left(\Delta I_i\right)\Delta I_i\right], \quad \forall i \in [1,n] \\ f(x) = \dfrac{x}{x^2 + k} \end{cases} \tag{1}$$

where $\{I_1,..., I_n\}$ is the noised video sequence, $\Delta I_i = \nabla^2 I_i$ and f represents a monotone decreasing noise smoothing function [13]. Other noise removal operations, such as median filtering and Wiener adaptive filtering, can be further applied on the PDE filtered image.

The denoised video is then normalized to enhance its contrast. Sharpening the contrast could help the image object detection process. We use the following normalization formula for the video sequence:

$$I_t^{'}(i,j) = 255 \cdot \frac{I_t(i,j) - \min(I_t)}{\max(I_t) - \min(I_t)}, \forall i \in [1,M],\ j \in [1,N], \tag{2}$$

where each frame I_t represents a $[M \times N]$ grayscale image.

3. Temporal differencing approach for moving object detection

Let the enhanced sonar movie sequence be noted as $Vid = \{I_1,...,I_n\}$, where I_i could represent here the video keyframes, to simplify the computation process. We propose an improved temporal differencing (frame-difference based) algorithm for moving object detection.

We assume that the moving video objects from *Vid* do not collide, are not very close to each other and do not stop until the they reach the last frame. The grayscale image obtained as difference of two successive frames indicates the video motion between them, its non-black regions representing the moving regions.

Unfortunately, a moving region does not represent always an entire moving object. An image object in a sonar frame can be composed of more homogeneous regions, characterized by various intensities. Some of them have higher intensities than that of the image background, while other regions have lower intensities. A filtered ultrasound image (frame) is characterized by a large and homogeneous background. We consider the following notation for the difference between two video frames:

$$DF\ (i,\ j) = I_i - I_j, \forall\ i \neq\ j \in [1, n]\,, \tag{3}$$

Any moving video object that is present in the frames I_i and I_j, is represented in image $DF(i, j)$ by some non-black regions corresponding to its homogeneous regions. The high-intensity regions are displayed in $DF(i, j)$ at the locations occupied in I_i, while low-intensity regions are displayed in $DF(i, j)$ at their positions in I_j.

Such a video frame difference example is displayed in Fig. 1. Two frames of an enhanced (denoised and normalized) sonar movie are represented in the figure, together with their difference image. One can see in that image the two lighter zones corresponding to a high intensity region and a low-intensity region of an object respectively.

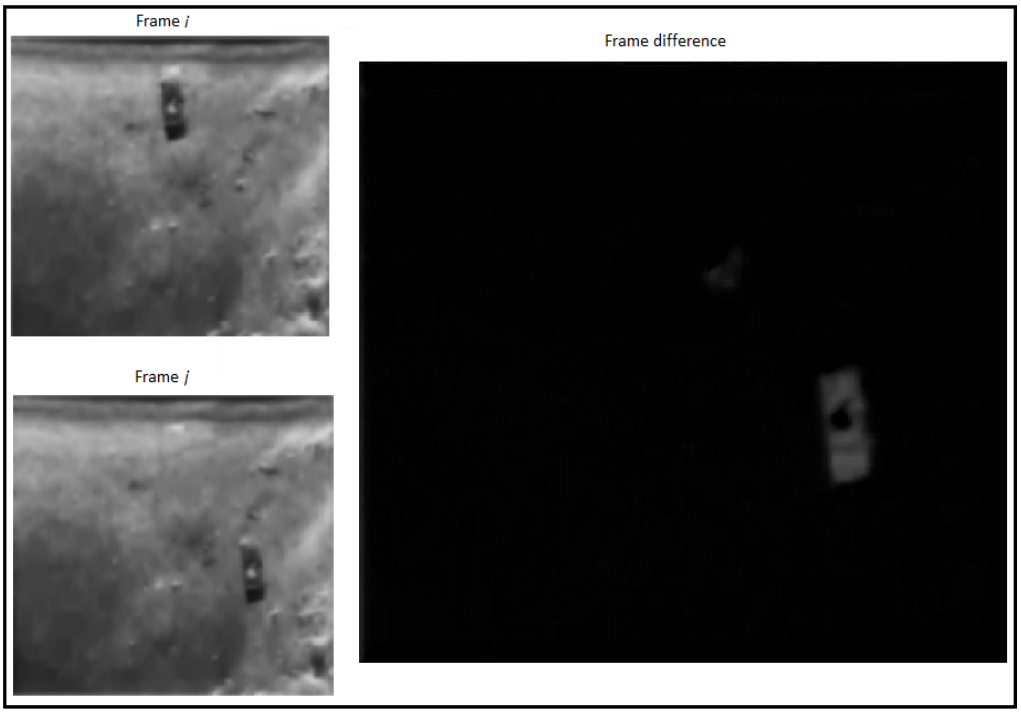

Fig. 1. Temporal differencing for two sonar frames

First, we propose an algorithm that determines the moving image objects from I_1 to I_{n-1}. Then, a special reasoning is used for the last frame.

At each step *i*, our algorithm detects the moving objects of the frame I_i, using the next video frames, I_{i+1} and I_{i+2}. It computes the video frame

differences $DF(i,i+1)$ and $DF(i,i+2)$ first. Then, these frame differences are converted into the binary form, by setting to 1 all their non-zero pixels (or all pixels exceeding a given threshold).

Let the resulted binary images be noted as $DF_B(i,i+1)$ and $DF_B(i,i+2)$ respectively. Next, some mathematical morphological operations are applied to them [16]. We consider a morphological closing process, representing a dilation followed by an erosion, to be performed on both the binary images. We obtain the following closed images:

$$\begin{cases} DF_B^c(i,i+1) = DF_B(i,i+1) \bullet D \\ DF_B^c(i,i+2) = DF_B(i,i+2) \bullet D \end{cases} \tag{4}$$

where $\bullet$ represents the closing operator and D is the chosen structuring element, representing a disk with radius 1. The connected components of the closed binary images are determined, the components having very small areas (under a selected threshold) being discarded.

The two morphologically processed frame difference images $DF_B^c(i,i+1)$ and $DF_B^c(i,i+2)$, corresponding to a video frame of a sonar movie sequence containing military vehicles, are described, as an example, in Fig. 2.

One computes the intersection of the binary images obtained from (4). Herein we consider the *intersection* of two binary images to be the image having the pixel values of the two binary images where they coincide and 0 in the locations where the two images differ. Therefore, the intersection image is obtained as:

$$Int_1(i) = DF_B^c(i,i+1) \bigcap DF_B^c(i,i+2) \tag{5}$$

which means

$$Int_1(i)[x,y] = \begin{cases} DF_B^c(i,i+1)[x,y] = DF_B^c(i,i+2)[x,y] \\ 0, DF_B^c(i,i+1)[x,y] \neq DF_B^c(i,i+2)[x,y] \end{cases} \tag{6}$$

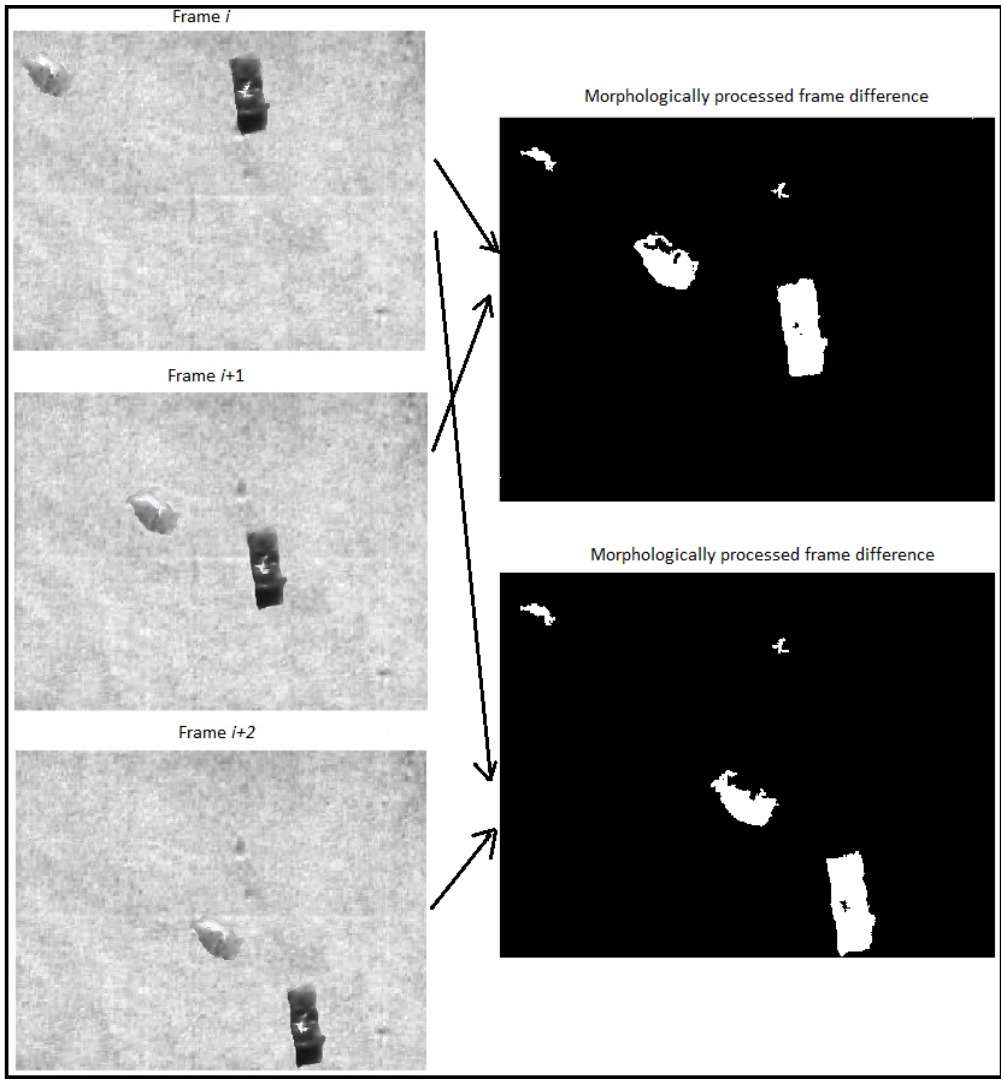

Fig. 2. Morphologically processed frame difference images example

The connected components of $Int_1(i)$ correspond to all high-intensity moving regions of I_i. Therefore, the high-intensity moving zones of the sonar frame I_i from Fig. 2 are those corresponding to the white regions from the next figure, obtained from (6).

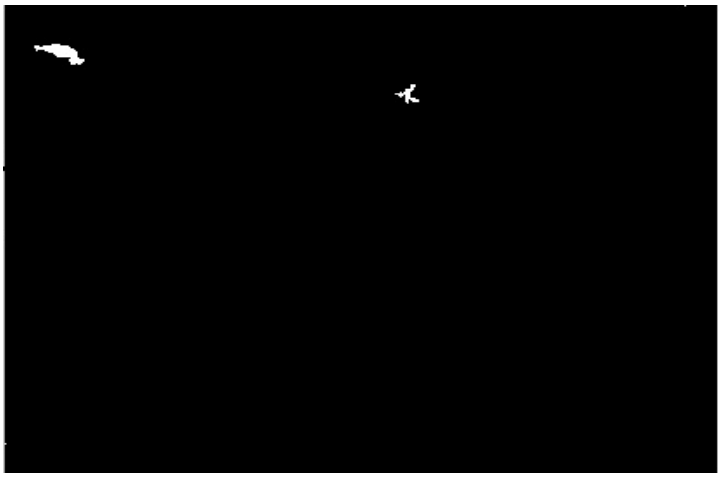

Fig. 3. Locations of the high-intensity regions of moving objects

The low-intensity moving regions of the video frame are determined using a similar identification algorithm. Thus, these regions correspond to the connected components of the binary intersection image

$$Int_2(i) = DF_B^c(i+1,i) \cap DF_B^c(i+2,i) \tag{7}$$

where the closed binary frame differences $DF_B^c(i+1,i)$ and $DF_B^c(i+2,i)$ are computed as in the previous case. Thus, the low-intensity regions of the frame I_i from Fig. 2 are those corresponding to the connected components of Fig. 4, obtained from relation (7).

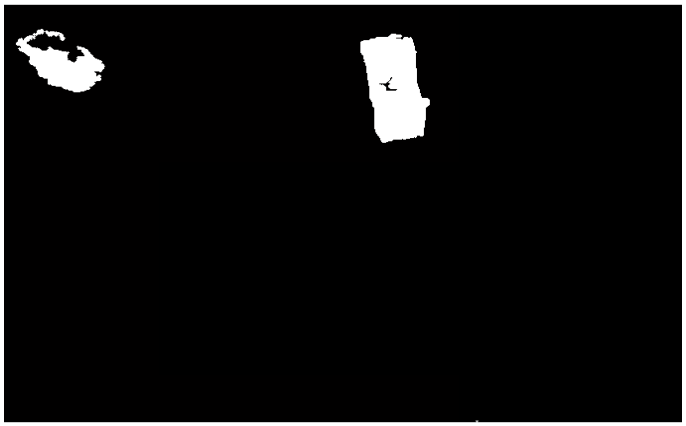

Fig. 4. Locations of the low-intensity regions of moving objects

The binary image representing the sum of these two image intersections is computed as:

$$Mov(i) = Int_1(i) + Int_2(i) \tag{8}$$

The connected components of *Mov*(*i*) correspond to the moving objects of the frame I_i. If the summing relation (8) is applied to binary images in Fig. 3 and Fig. 4, the image displayed in Fig. 5 is obtained. The two connected components represent the locations of the moving objects of the first sonar video frame from Fig. 2.

Fig. 5. Locations of the moving objects of I_i

For $I_1,...,I_{n-1}$ the moving video objects are detected in the same way. The last step of the detection technique consists of identification of the final locations in I_n of the moving objects. Let us name these zones of I_n as *moving objects* of the frame, although this term somewhat incorrect here. Their detection is performed through a backward process given by the relation $Mov(n) = Int_1(n) + Int_2(n)$, where

$$Int_1(n) = DF_B^c(n, n-1) \bigcap DF_B^c(n, n-2) \tag{9}$$

and

$$Int_2(n) = DF_B^c(n-1,n) \bigcap DF_B^c(n-2,n) \tag{10}$$

As the result of the proposed procedure, there are *K* detected image objects in each video frame. The next step is the tracking of these objects in the video sequence.

4. Video object tracking using a texture analysis approach

In this section we propose a video tracking technique for the detected moving objects. So, we consider an object matching approach using the texture analysis of the *K* image objects identified in each frame.

One knows the moving image objects of each frame, but the instances of any given object in the next frames remain unknown. Our video tracking approach finds for each image object of a frame its instance in the next frame, representing the correspondent moving image object. In our approach the matching object has to be the most similar object from the next frame.

Therefore, the considered tracking method is based on an image object recognition process. The first step of this recognition process is a content-based feature extraction for image objects. The second step consists of a supervised feature vector classification based on a minimum distance classifier.

We note the sequence of the moving image objects from the video frame I_i as:

$$Ob(i) = \left\{Ob_1^i,..., Ob_K^i\right\}, i \in [1,n] \tag{11}$$

these objects being numbered in the image from left to right and from up to down. Any tracked video object is modeled as an ordered sequence of similar image objects, $\left[Ob_j^1,...,Ob_{pos(j,i)}^i,...,Ob_{pos(j,n)}^n\right]$, where the content similarity is expressed as:

$$Ob_j^1 \approx ... \approx Ob_{pos(j,i)}^i \approx ... \approx Ob_{pos(j,n)}^n \tag{12}$$

where $pos(j,i) \in [1,K]$. We do not consider a shape similarity between these objects, because the sonar image objects may often have similar shapes. We consider a texture-based content similarity for video analysis instead.

For each moving object Ob_j^i one determines the sub-image $Im(Ob_j^i)$ corresponding to its bounding box. A texture analysis is performed for each subimage. Thus, we propose a texture-based content descriptor using 2D Gabor filters for the analyzed sub image [15].

Two-dimension Gabor filtering represents a widely used tool in image texture analysis and many other image processing and analysis domains [15,17]. The Gabor filter constitutes a band-pass linear filter whose impulse response is defined by a harmonic function multiplied by a Gaussian function.

The two-dimensional Gabor filter can be viewed as a sinusoidal plane of particular frequency and orientation, modulated by a Gaussian envelope. We consider an even-symmetric 2D Gabor filter, having the following form:

$$G_{\theta_k,f_i,\sigma_x,\sigma_y}(x,y)=\exp\left(-\left[\frac{x_{\theta_k}^2}{\sigma_x^2}+\frac{y_{\theta_k}^2}{\sigma_y^2}\right]\right)\cdot\cos\left(2\pi f_i x_{\theta_k}\right) \tag{13}$$

where $x_{\theta_k}=x\sin\theta_k+y\cos\theta_k$, $y_{\theta_k}=x\cos\theta_k-y\sin\theta_k$, f_i is the central frequency of the sinusoidal plane wave at the angle $\theta_k=\frac{k\pi}{n}$ with the $x-$axis, σ_x and σ_y are the standard deviations of the Gaussian envelope along the two axes [15].

The image of each object Ob_j^i is filtered with $G_{\theta_k,f_i,\sigma_x,\sigma_y}$ at various orientations, radial frequencies and standard deviations. We select some proper filter parameters: $\sigma_x=2$, $\sigma_y=1$, $f_i\in\{0.75,1.5\}$ and $\theta_k\in\left\{\frac{\pi}{3},\frac{2\pi}{3},\pi\right\}$.

The resulted 2D Gabor filter bank $\left\{G_{\theta_k,f_i,2,1}\right\}_{f_i\in\{0.75,1.5\},k\in[1,3]}$, composed of 6 channels, is applied to the object image, by convolving it with each Gabor filter from the set. The resulted Gabor responses are concatenated into a 3D feature vector. The feature extraction process is modelled as following:

$$V(Ob_j^i)[x,y,z]=V_{\theta(z),f(z),\sigma_x,\sigma_y}(Ob_j^i)[x,y] \tag{14}$$

where

$$\theta(z)=\begin{cases}\theta_z, & z\in[1,3]\\ \theta_{z-n}, & z\in[4,6]\end{cases},\quad f(z)=\begin{cases}f_1, & z\in[1,3]\\ f_2, & z\in[4,6]\end{cases} \tag{15}$$

and

$$V_{\theta(z),f(z),\sigma_x,\sigma_y}(Im(Ob_j^i))[x,y]=Im(Ob_j^i)(x,y)\otimes G_{\theta(z),f(z),\sigma_x,\sigma_y}(x,y) \tag{16}$$

So, for each moving object we get a 3D feature vector $V(Ob_j^i)$, that is a robust content descriptor. The moving objects are compared by computing the Euclidian distance between their feature vectors. Obviously, the analyzed objects may have various dimensions, so a resizing process must be performed on their images before applying the Euclidian metric.

The video object tracking process has $n - 1$ steps. At each step i, the matching algorithm detects the instances of all objects Ob_j^i in the next frame, I_{i+1}. The next instance of any video object is the most relevant object of I_{i+1}, that is the image object corresponding to the minimum distance between feature vectors. Thus, one identifies the matching object $Ob_{ind}^{i+1} \approx Ob_j^i$, where

$$ind = \arg\min_{k\in[1,K]} d(V(Ob_j^i), V(Ob_k^{i+1})), \forall i \in [1, n-1],\ j \in [1, K] \tag{17}$$

where d represents the Euclidean distance.

An example based on this video object tracking process is represented in Fig. 6. Object matching is applied on the same sequence of the three consecutive video frames analyzed for object detection.

The detected objects, two for each frame, are bounded by colored rectangles. The identified match of each object is marked by the same color (red or blue). The arrows indicate the trajectory of the tracked video object.

5. Experiments

The video object detection and tracking techniques proposed in this paper have been tested on various sonar movie datasets. We have performed numerous video analysis experiments on ultrasound image sequences and obtained satisfactory results, which prove the effectiveness of our techniques.

High detection and tracking rates are produced by the provided approaches. The moving object identification method has a detection rate of approximately 90%.

There have been obtained high values for performance parameters of *Precision*, *Recall* and F_1. That means there are very few missed hits (undetected moving objects) and wrongly detected image objects.

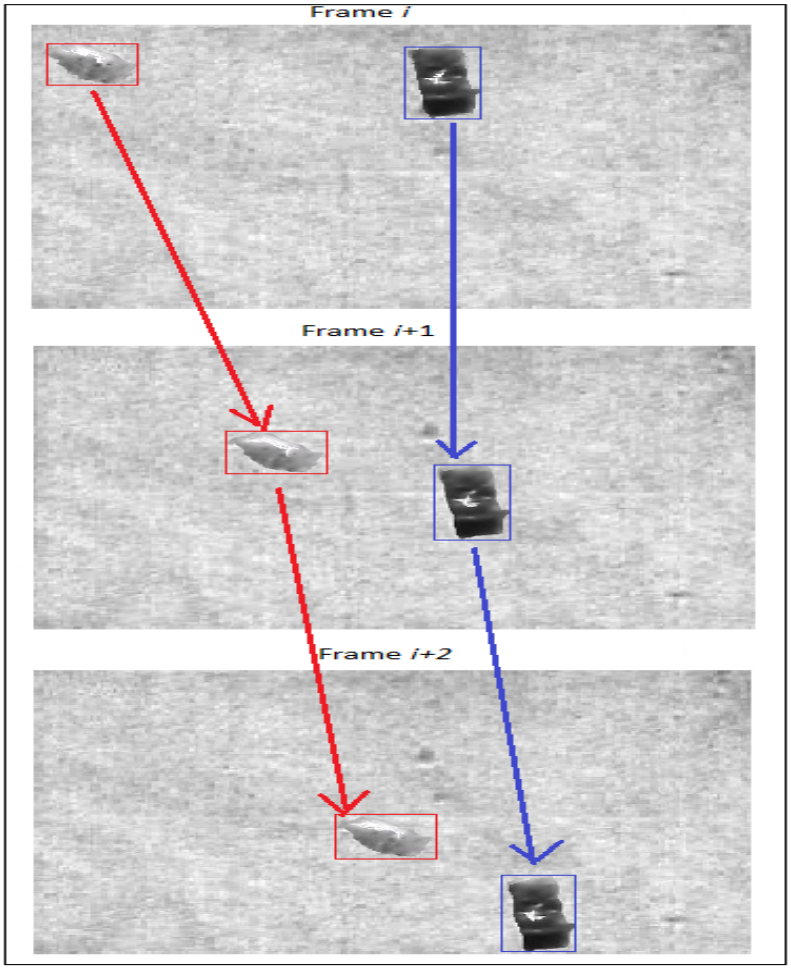

Fig. 6. Object matching process in a sonar video

The video tracking algorithm is characterized by a high object recognition rate that is over 80%. The performance parameters are *Precision* = 0.80, *Recall* = 0.85. This means there is a small number of false positives and false negatives too.

6. Conclusions

We have developed an automatic video object detection and tracking system for ultrasound movies, characterized by a fixed camera and the presence of

multiple moving objects. This paper brings important contributions in both the object detection and video object tracking fields.

Thus, we have provided a novel moving object detection technique. Our detection approach uses an improved frame difference algorithm to detect the video motion. This temporal differencing method uses the differences between the current frame and the next two frames, and some morphological operations applied to the binarized frame differences.

In the video tracking stage we have described a novel image object matching method. The proposed object recognition approach uses a 2D Gabor filtering based feature extraction, that produces robust object content descriptors, and a supervised minimum distance based object classification approach.

The frame difference based technique and the texture analysis method represent the main original contributions of this article. The automatic character of our detection and tracking system is also a very important thing, no interactivity being required.

The detection and tracking results described here can be successfully applied in numerous important domains such as video indexing and retrieval, robotics, sonar navigation and video surveillance. Our future research in this video analysis domain will focus on improving the proposed techniques. Therefore, we intend to make them work for colliding video objects or some camera motions.

Acknowledgments

The research described here has been supported by the grant PN II, Programme 4 - Partnerships in priorities domains 2007-2013, Project type: PC – Complex projects**, Adbiosonar – adaptive bio-mimetic sonar heads for autonomous vehicles**, Contract no: 12079/2008.

R E F E R E N C E S

[1] *A. Yilmaz*, "Object Tracking: A Survey", ACM Computing Surveys, Vol. 38, No. 4, Article 13, December 2006.

[2] *J. Shi, J. Malik*, "Normalized cuts and image segmentation", IEEE Trans. Patt. Analy. Mach., Intell. *22,* 8, pp. 888–905, 2000.

[3] *T. Barbu*, "A Pattern Recognition Approach to Image Segmentation", Proceedings of the Romanian Academy, Series A, Volume 4, Number 2, pp.143-148, May-August 2003.

[4] *C. Wren, A. Azarbajeani, A. Pentland*, "Pfinder: Real-time tracking of the human body", IEEE Trans. Patt. Analy. Mach. Intell. 19, 7, pp. 780–785, 1997.

[5] *R. Jain, H. Nagel*, "On the analysis of accumulative difference pictures from image sequences of real world scenes", IEEE Trans. Patt. Analy. Mach. Intell. 1, 2, pp. 206–214, 1979.

[6] *M. Kass, A. Witkin, D. Terzopoulos*, "Snakes: Active Contour Models", International Journal of Computer Vision, pp. 321-331, 1988.
[7] *P. Lappas, J. N. Carter, R. I. Damper*, "Robust evidence-based object tracking", Journal Pattern Recognition Letters, Volume 23 Issue 1-3, January 2002, Elsevier Science Inc. New York, NY, USA.
[8] *X. Hu, Y. Tang, Z. Zhang*, "Video Object Matching based on SIFT Algorithm", in Proceedings of International Conference on Neural Networks and Signal Processing, pp. 412-415, 2008.
[9] *N. Peterfreund,* "Robust Tracking of Position and Velocity With Kalman Snakes", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 21, no. 6, June 1999.
[10] *D. Comaniciu*, "Kernel-based Object Tracking", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25, no. 5, 2003.
[11] *L. Wixson*, "Detecting Salient Motion by Accumulating Directionally-Consistent Flow", IEEE Transactions on pattern analysis and machine intelligence, vol. 22, no. 8, August 2000.
[12] *E. Trucco, Y. R. Petillot, I. Tena Ruiz, K. Plakas, D. M. Lane*, "Feature Tracking in Video and Sonar Subsea Sequences with Applications", Computer Vision and Image Understanding, 79**,** pp. 92–122, 2000.
[13] *T. Barbu, V. Barbu, V. Biga, D. Coca*, "A PDE variational approach to image denoising and restoration", Nonlinear Analysis: Real World Applications, Volume 10, Issue 3, pp. 1351-1361, June 2009.
[14] *T. Barbu*, "A PDE based Model for Sonar Image and Video Denoising", Analele Stiintifice ale Universitatii Ovidius, Constanta, Seria Matematica, Volumul 19, Fascicola 2, 2011.
[15] *T. Barbu*, "Content-based image retrieval system using Gabor filtering", Proceedings of the 20th International Workshop on Database and Expert Systems Applications Mining and Management, pp. 236-240, 31 August - 4 September, Linz, Austria, 2009.
[16] *J. Serra, P. Soille*, "Mathematical Morphology and Its Applications to Image Processing", Proceedings of the 2nd international symposium on mathematical morphology (ISMM'94), 1994.
[17] *A. Jain, N. Ratha, S. Lakshmanan*, "Object detection using Gabor filters", Pattern Recognition, vol. 30, pp. 295-309, 1997.