International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:1, No:12, 2007

An Optical Flow Based Segmentation Method for Objects Extraction

C. Lodato, and S. Lopes

***Abstract*—**This paper describes a segmentation algorithm based on the cooperation of an optical flow estimation method with edge detection and region growing procedures.

The proposed method has been developed as a pre-processing stage to be used in methodologies and tools for video/image indexing and retrieval by content. The addressed problem consists in extracting whole objects from background for producing images of single complete objects from videos or photos. The extracted images are used for calculating the object visual features necessary for both indexing and retrieval processes.

The first task of the algorithm exploits the cues from motion analysis for moving area detection. Objects and background are then refined using respectively edge detection and region growing procedures. These tasks are iteratively performed until objects and background are completely resolved.

The developed method has been applied to a variety of indoor and outdoor scenes where objects of different type and shape are represented on variously textured background.

***Keywords*—**Motion Detection, Object Extraction, Optical Flow, Segmentation.

I. Introduction

A fundamental step for video/image retrieval by content is the calculation of the visual features. In some applicative contexts the most relevant content consists in represented subjects rather than the whole scene. For example, let prefigure an application that takes in input a photo or a video sequence, extracts the focused objects, searches similar images in its database and outputs the corresponding descriptions.

In this perspective, the distinction between focused objects and background permits to calculate the visual features of each object alone, making more effective the retrieval task. To this end, a segmentation method for the extraction of images of single complete objects from a digital photo or a short video sequences has been developed.

Image segmentation, widely employed in medical imaging, computer vision, production quality control, etc [1]-[3], is the process to extract meaningful regions from an image. However, the definition of meaningful region is strictly dependent from the applicative context. Generally, an image segmentation process should capture image parts that are perceptually relevant. The definition of what is perceptual relevant characterizes any segmentation method. Anyway, the resulting segmentation should produce a collection of some global high-level characteristics of the image. In this paper, the meaningful regions are these related to the focused objects. Thus, the goal is to produce images including only a single object on a conventional background, free of any other particular that could influence the next retrieval process. Many image segmentation techniques can be found in literature. They can be roughly classified in region-based and contour-based approaches. Region-based approaches try to find sub-sets of the image pixels characterized by coherent visual properties such as brightness, colour and texture [4], [5]. Contour-based approaches try to generate closed boundary images. Some methods are based on a preliminary step of edge detection and a process of connecting edges together in order to get extended contours [6]-[9]. The edge detection step uses a discriminating threshold value. If this value is too low many random edges will be detected, on the contrary some edges will be missed. More recently, methods based on the active contour model have been proposed. They try to fit an initial boundary curve or shape to the contours of objects [10]-[12]. A further evolution is represented by geodesic contour model in which the initial parameter is represented by one or more points belonging to the object to be segmented [13]. It should be also considered that many images can not be segmented using only local decision criteria in either region-based or contour-based approaches. Some kind of adaptive or non local criterion must be used in order to capture global aspects of the image. In any case, the segmentation can produce too many regions (over segmentation) or too few regions (under segmentation), even in the same image. For what concerns the need of distinguishing whole objects from the background, the segmentation process should be able to gather all the internal structures of a single object together and, at the same time, discard or even cancel eventual structures belonging to the background. In this sense, the segmentation process should not be either too coarse, or too fine.

C. Lodato is with the Istituto di CAlcolo e Reti ad alte prestazioni (ICAR) of Consiglio Nazionale delle Ricerche, viale delle Scienze, edificio 11, 90128 Palermo, Italy, phone: +39-091-238-111; fax: +39-091-652-9124; e-mail: c.lodato@icar.cnr.it.

S. Lopes is with the Istituto di CAlcolo e Reti ad alte prestazioni (ICAR) of Consiglio Nazionale delle Ricerche, viale delle Scienze, edificio 11, 90128 Palermo, Italy, phone: +39-091-238-111; fax: +39-091-652-9124; e-mail: s.lopes@icar.cnr.it.

II. Method Description

The necessity to detect the object as a whole and not only

part of it has been approached using a hybrid segmentation technique based on edge detection, region growing, and optical flow procedures. The use of optical flow procedures together with other segmentation techniques has been already exploited in other works [14]-[19], but to the end to get a more accurate estimation of the motion field. If the background is plain, the focused object is easily detected as a whole by means of optical flow estimation in the real scene (video sequence) or in the synthetic scene built applying a shift to the image (photo). This is a trivial task when the background is characterized by a perfect uniform brightness and it can easily performed also by the above mentioned techniques. However, this ideal requirement is rarely fulfilled because noise, shadows, corners, texture or other discontinuities in the background could result in detected moving areas that do not match with the focused object. In this perspective, the optical flow estimation could not generally supply enough information for discriminating the object motion from that of the background. Thus, it is necessary to adopt a strategy for reducing the influence of these factors. Precisely, an edge detection method is used for locating that points that present a brightness discontinuity, as in correspondence to the edge between the focused object and background. However, brightness discontinuities could also be present in some part of the background or inside the object. For this reason, an object map, representing the pixels the object is made of, is derived by the comparison between the motion field and the edge map.

The image area belonging to the background is detected applying a region growing procedure using as seeds the pixels in which the motion field is negligible. In correspondence to the region map points the motion field is then set to zero and the brightness of the images are set to the same value, producing a new image sequence. Moreover, the edge map is upgraded removing those edges lying on the region map. The above actions are iteratively repeated until the object and the background are completely resolved.

A. The Algorithm

The algorithm is made of cooperating procedures that perform four basic tasks: optical flow estimation, moving object detection, edge segmentation and region growing. These procedures are iteratively applied in order to refine the image of the extracted objects.

The procedure developed for the optical flow estimation belongs to the class of differential methods [20] in the sense that it exploits the spatial-temporal variations of brightness. It is an improved version of the optical flow estimator reported in [21]. Differential methods are generally based on the following equation which should hold true for every point of the images:

$$I_{t_1}(x, y) \approx I_{t_0}(x - udt, y - vdt) \tag{1}$$

where $I_t(x, y)$ represents the brightness of a generic point p located at coordinates (x, y) at time t, whereas u and v represent the unknown velocity components of p. Thus, it is necessary to impose other constraints in order to resolve the problem *(aperture problem)* [22]. To this end, let consider the square lattice derived skipping, for example, all even rows and columns of the image starting from the pixel at location (i_0, j_0) in Fig. 1. The motion values corresponding to this pixel subset, marked as black dots, are considered as unknown. The motion values of all the other pixels, the grey dots, can be expressed interpolating the values corresponding to the neighbours belonging to the lattice. The interpolation relations represent the additional constraints considered in order to make the problem tractable.

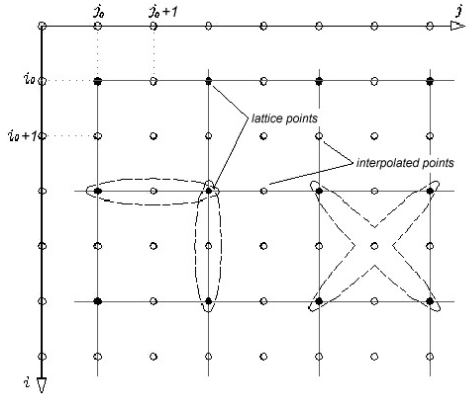


Fig. 1 Lattice scheme

The motion field is thus derived minimising the functional that express the brightness difference between the image I_1^*, built applying the motion field to the first image of the sequence, and the second image:

$$E = \sum_{i=0}^{i=m-1} \sum_{j=0}^{j=n-1} \left(I_1^*(i, j) - I_1(i, j) \right)^2 , \tag{2}$$

where *m*, *n* are the vertical and horizontal sizes of the images expressed in pixels. Let $Sv_{i,j}$ and $So_{i,j}$ respectively be the motions along the vertical and the horizontal axis related to the pixel of coordinate (i, j), the brightness intensity $I_1^*(i, j)$ can be expressed by the following equation:

$$\begin{aligned} I_1^*(i, j) \;=\; & I_0(i_s, j_s) \cdot (1 - |\, sv_{i,j}\,|) \cdot (1 - |\, so_{i,j}\,|) + \\ & I_0(i_s, j_s - vo_{i,j}) \cdot (1 - |\, sv_{i,j}\,|) \cdot |\, so_{i,j}\,| + \\ & I_0(i_s - vv_{i,j}, j_s) \cdot (1 - |\, so_{i,j}\,|) \cdot |\, sv_{i,j}\,| + \\ & I_0(i_s - vv_{i,j}, j_s - vo_{i,j}) \cdot |\, so_{i,j}\,| \cdot |\, sv_{i,j}\,|, \end{aligned} \tag{3}$$

with

$$i_s = i - [Sv_{i,j}], \quad sv_{i,j} = Sv_{i,j} - [Sv_{i,j}], \quad vv_{i,j} = sv_{i,j} / |\, sv_{i,j}\,|,$$
$$j_s = j - [So_{i,j}], \quad so_{i,j} = So_{i,j} - [So_{i,j}], \quad vo_{i,j} = so_{i,j} / |\, so_{i,j}\,|.$$

Equation (3) could be used for calculating the brightness variation caused by any magnitude of movement, but it is no longer valid if the distance between pixel (i, j) and the edge of the object, which is opposite the direction of movement, is less than the shift. Defining as $err(i, j)$ the difference between the calculated brightness $I_1^*(i, j)$ and the real one it follows that:

$$err(i, j) = I_1^*(i, j) - I_1(i, j) . \quad (4)$$

Combining (2) with (3) yields:

$$err(i, j) = a_{i,j} \cdot | sv_{i,j} | \cdot | so_{i,j} | + b_{i,j} \cdot | sv_{i,j} | + c_{i,j} \cdot | so_{i,j} | + d_{i,j}, \quad (5)$$

where

$$a_{i,j} = I_0(i_s, j_s) + I_0(i_s - vv_{i,j}, j_s - vo_{i,j}) + - I_0(i_s, j_s - vo_{i,j}) - I_0(i_s - vv_{i,j}, j_s)$$
$$b_{i,j} = I_0(i_s - vv_{i,j}, j_s) - I_0(i_s, j_s)$$
$$c_{i,j} = I_0(i_s, j_s - vo_{i,j}) - I_0(i_s, j_s)$$
$$d_{i,j} = I_0(i_s, j_s) - I_1(i, j) .$$

The coefficients $a_{i,j}, b_{i,j}, c_{i,j}, d_{i,j}$ depend on the brightness value of the pixels involved in the motion. The second order term plays a particularly important role as (5) is still useful in cases where the brightness gradient is null. The optical flow is obtained by minimising the following functional:

$$E = \sum_{i=0}^{i=m-1} \sum_{j=0}^{j=n-1} err(i, j)^2 . \quad (6)$$

The minimisation process of the above functional is performed with an iterative algorithm described in [21]. In order to make equal the contribution of every pixel to the flow estimation, each pixel should be considered at least once as point of the lattice. This can be done by using four different lattice positions, at coordinates (i_0, j_0) $(i_0, j_0 + 1)$ $(i_0 + 1, j_0 + 1)$ $(i_0 + 1, j_0)$. Hence, the optical flow determination results from the four independent minimisation problems equivalent to the original one. They differ from each other in the lattice position only. The achieved fields are smoothed and averaged, yielding a single motion field. It should be noticed that for the purpose of the segmentation, is not useful to get an accurate motion field at every step. Hence, the minimisation process is stopped after few iteration. The main reason for the early stopping the optical flow estimation procedure is that the real motion could regard the whole image. The preliminary estimation of motion will produce significant values of velocity in correspondence to points characterised by a higher brightness gradient. Hence, when an object stands out against the background, its contour is surely characterised by a significant motion. For the reason explained in the follows, the processed image sequence is upgraded every time the optical flow estimation procedure restarts. The resulting motion field is used as input for the object detection procedure.

The next stage of the presented method is devoted to the detection of motion field zones which can be considered as moving objects. This task is accomplished aggregating adjacent pixels characterised by non null values of the velocity that are congruent with a rigid motion. These zones are labelled in an object map and thus refined using information derived from the edge detection procedure. This procedure computes the brightness gradient of each pixel with his neighbours. If it is greater than a threshold value, the pixel is marked as edge, producing an edge map EM. Let $\mathbf{G}_x(i, j) = I(i, j) - I(i, j-1)$ and $\mathbf{G}_y(i, j) = I(i, j) - I(i-1, j)$ the components of the brightness gradient vector $\mathbf{G}$, EM is a binary image built with the following rule:

$$EM(i, j) = 1 \quad \text{if} \quad |\mathbf{G}| \geq T$$
$$EM(i, j) = 0 \quad \text{otherwise,}$$

where T is the threshold value. The threshold value has generally a crucial influence on the results of all edge detection procedures for image segmentation and there are many methods for the calculation of a suitable value [23], [24]. It is assumed that the brightness differences across the boundary between the object and background are generally higher than many of the differences within the background. Hence, a suitable value can be calculated in such a way to let a small number of still detectable edges on the background. Since at the beginning the background is not defined yet, the initial threshold value is thus calculated off cycle and results from the brightness difference distribution of pixels located on a narrow border strip of the image. This choice is due to the assumption that the focused object is almost completely surrounded by the background. The value is calculated setting that no more than a fixed percentage of the considered pixels (typically 0.1%) have a brightness difference exceeding that value. In any case, the threshold value can not be lower than 6, because this value is the lowest reasonably brightness difference perceptible by a human eye for 256 grey-levels images. The first edge map EM is built using the initial threshold value and a first boundary map is derived marking all the pixel of the edge map between the first and the last detected edge for every row and column. The unmarked areas of the boundary map represent a first evaluation of regions that likely belong to the background, and are used for evaluating a new threshold value that results from a more significant pixel set. The edge map is employed to better define the contour of the objects detected by the moving object detection procedure.

A new motion field is produced assigning to all the pixels of each object motion values congruent with its corresponding

rigid motion. Further information for improving on the background definition are obtained applying a region growing technique to the image.

Many techniques of region growing for image segmentation can be found in literature [25]-[29]. They generally differ from each other in the similarity criterion. A weak point of seeded region growing procedure is the choice of the number and location of seed points. In the presented method, information derived by the optical flow procedure provide a reasonable criterion for this choice. In fact, the seeds are located in correspondence to the pixels that not belong to moving objects and, at the same time, have negligible motion. The similarity criterion considered is based on the pixel brightness. When the brightness difference between a pixel not belonging to a moving object and an adjacent region is less than a threshold value, the pixel is aggregated to the region. Region growing and edge detection procedures have the same threshold value. Since the regions grow from seeds located in pixels where the motion field is either null or negligible, the regions should progressively cover the background. Hence, the motion field is set to zero in correspondence to the pixels belonging to the current region map. Furthermore, the region map is also used for recalculating a more effective threshold value based on a progressively wider and more significant data domain. That is, the new threshold value is obtained from the brightness difference distribution of the image pixels located on the region map. The resulting motion field is compared with the one calculated at the previous iteration cycle. If the value of the functional expressed by (6) computed on the motion field at the generic step k is lower than the previous one, that is $E_k < E_{k-1}$, the new field is accepted and the unmarked area of the region map represents an intermediate solution to the problem. In the latter case, the brightness value of the pixels of the second image I_1 belonging to the region map is set equal to the correspondent one of the first image I_0. In this way, the next image sequence to be processed is approximated to the case where the background is still. Hence, the algorithm restarts a new iteration cycle using the accepted motion field as initial point for the optical flow estimation. The algorithm stops when the decrease of the functional E is less than a prefixed value.

To exploits information provided by the optical flow estimation for image segmentation, is motivated by the assumption that an object stand out against a plain background. The adopted strategy that merges also the cues provided by edge detection and region growing procedures, permits to obtain good results also in other situation as it is shown in the following section.

III. Experimental Results

A set of experiments has been carried out in order to test the effectiveness of the presented method. Three images from the Columbia COIL image database are shown in Fig. 2(a). A test on these images is reported in [4] but with a different aim. They are a good example for the purpose of the presented method that is to extract the whole object. The objects are shown on an apparently plain background and are characterised by a substantial intensity gradient across their faces. Moreover, some of their boundaries have brightness comparable with the background. Although the edges of the bottom of the cup and the right side of the pot are almost imperceptible, the extracted objects are well defined as can be noticed from the images in Fig. 2(b).

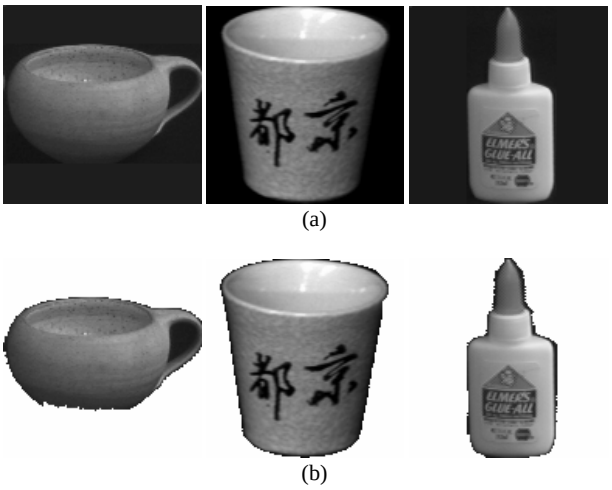


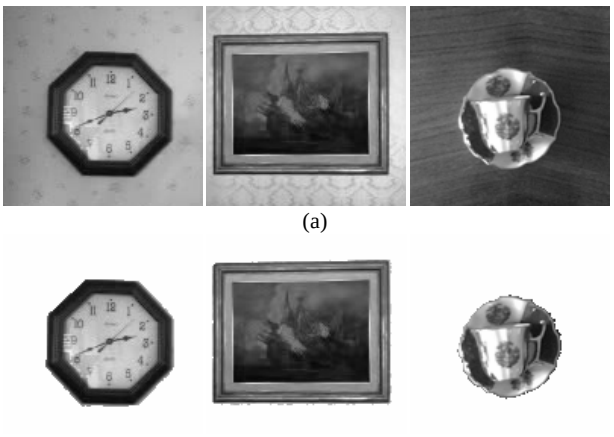
Fig. 2 (a) Three images from the COIL data base, (b) extracted objects

The following sequences have been acquired by means of a common mobile videophone, with a resolution of 144x144 pixels, but the developed method do not pose any constraint on the image size. From each sequence, about two seconds long, two images have been extracted. All the sequences have been shot indoor, with both natural and electric light. The reported cases differ in terms of number, shape and size of the focused objects, and of background texture.

In the first set of scenes (*wall clock*, *painting*, and *cup*) only a single object is focused as it can be seen in Fig. 3(a). The background is characterised by different textures. Precisely, there are two types of wallpaper background for *wall clock* and *painting* and the scenes have been filmed with natural light. The coffee cup is on a wooden table and the scene has been filmed with electric light coming from the left. In this case, the background is characterised by a wood grain texture. The images of the extracted object are shown in Fig. 3(b). The conventional background has been represented in white. As can be noticed the segmentation is well performed. The contour of the extracted object does not appear perfectly regular only in those parts where the boundary is not sharp.

In the second set (*statue*, *paintings*, and *books*) more objects are focused as it can be seen in Fig. 4(a). The background is of wallpaper type in the first two cases (natural light) whereas a wood grain texture characterises the background of the last example (electric light). All the objects

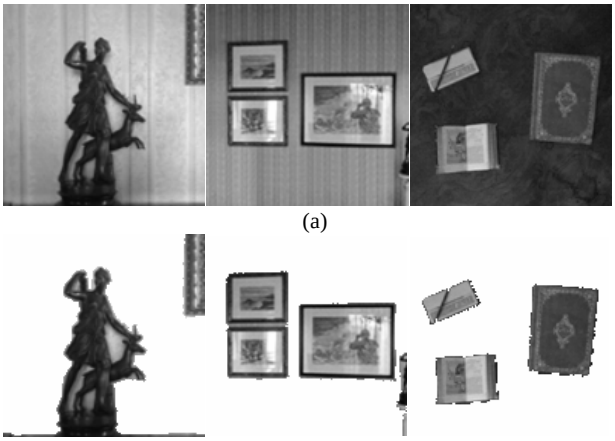
extracted are grouped in a single image for sake of simplicity, as it can be seen in Fig. 4(b).



(a)

(b)

Fig. 3 (a) Single object on variously textured background, (b) objects extracted with the proposed method



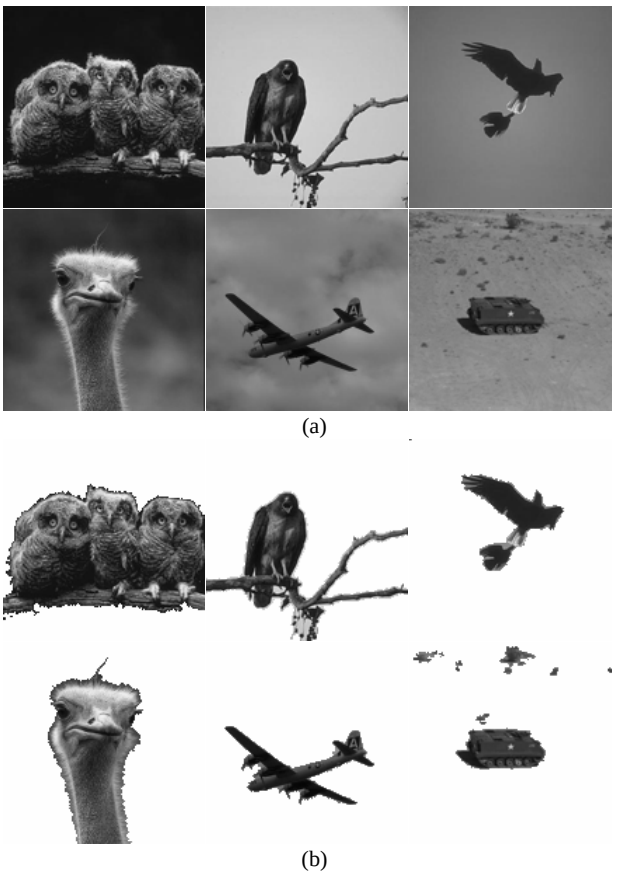
(a)

(a)

Fig. 4 (a) Multiple objects on variously textured background, (b) objects extracted with the proposed method

In *statue*, the scene also includes a portion of a picture frame and both objects are correctly extracted. The small portions of background in the statue are not errors. In fact, as already said in section II, information produced by optical flow estimation are used to group all the parts an object is made of. So that, any portion of background completely surrounded by an object is considered belonging to the object itself. The region growing procedure, devoted to the background detection, does not allow regions to expand themselves inside the objects. Anyway, the retrieval task is not affected because the image indexing is performed using images extracted with the same method.

The segmentation is well performed also when there are multiple focused objects in the scene, as can be seen for *paintings* and *books*. The two smaller paintings are detected as a single object because they are too close each other, whereas the small statue on the right lower part has been isolated from the nearby painting. All the objects in *books* are correctly resolved. The irregularities in some parts of the books contour are due to the brightness similarity between their covers and the table.



(a)

(b)

Fig. 6 (a) Outdoor scenes from the Berkeley Segmentation Dataset and the USC-SIPI Image Database, (b) extracted items

The method achieves satisfactory results also in outdoor scenes, provided that the background is not highly contrasted and the objects stand out against it. Some images from the Berkeley Segmentation Dataset and the USC-SIPI Image Database are reported in Fig. 6(a). The results can be considered more than satisfactory in particular for the three images in the bottom row where the background brightness is not uniform.

IV. Conclusion

In this paper a method for the automatic object extraction from a photo or a short video sequence of focused objects has been described.

The presented strategy is based on the cooperation of optical flow estimation, edge detection and region growing

algorithms.

Several tests have been carried out for verifying the method performances. Some of the reported exampled consists in scenes that have been filmed with a common videophone, with natural or electric light. The other included tests have been carried out on images from the Columbia COIL image database, the Berkeley Segmentation Dataset and the USC-SIPI Image Database.

No smoothing procedures or filters have been applied to reduce the noise in the processed images. The included tests results show the effectiveness of the method for different background types, shape and sizes of the objects. Good results have been achieved also in the case of multiple objects and variously textured background.

This research has been conducted as part of a wider project about methodologies and tools for video/image indexing and retrieval by content.

REFERENCES

[1] R. Pohle and K. Toennies, "Segmentation of medical images using adaptive region growing", *Proc. SPIE Medical Imaging 2001*, San Diego, CA, 2001, 1337-1346.

[2] G. A. Ruza, and P. A. Estéveza, "Image segmentation using fuzzy min-max neural networks for wood defect detection", *Proc. IPROMS 2005*, online web-based conference, July 2005, to be published by Elsevier.

[3] A. K. Jain and M. P. Dubuisson, "Segmentation of x-ray and c-scan images of fiber reinforced composite materials", *Pattern Recognition*, 25, pp. 257-269, 1992.

[4] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph based image segmentation", *International Journal of Computer Vision* 59(2), 167-181, 2004.

[5] J. Malik, S. Belongie, T. Leung and J. Shi, "Contour and texture analysis for image segmentation", *International Journal of Computer Vision* 43(1), 7-27, 2001

[6] U. Montanari, "On the optimal detection of curves in noisy pictures", *Comm. of the ACM*, vol.14:335-345, 1971.

[7] P. Parent and S. Zucker, "Trace inference, curvature, consistency, and curve detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence* Volume 11 , Issue 8, Pages: 823 – 839, August 1989.

[8] A. Sha'ashua and S. Ullman, "Structural saliency: the detection of globally salient structures using a locally connected network", in *Proc. 2nd Int. Conf. Computer Vision*, Tampa, FL, USA, 1988, pp 321-327.

[9] L. Williams and D. Jacobs, "Stochastic completion fields: a neural model of illusory contour shape and salience", in *Proc. 5th Int. Conf. Computer Vision*, Cambridge, MA, 1995, pp. 408-415.

[10] L. A. Vese and T. F. Chan, "A multiphase level set framework for image segmentation using the Mumford and Shah model", *International Journal of Computer Vision* 50(3), 271-293, 2002.

[11] N. Paragios, "A variational approach for the segmentation of the left ventricle in cardiac image analysis", *International Journal of Computer Vision* 50(3), 345-362, 2002.

[12] X. Wang, L. He and W. Wee, "Deformable contour method: a constrained optimization approach", *International Journal of Computer Vision* 59(1), 87-108, 2004.

[13] B. Appleton and H. Talbot, "Globally optimal geodesic active contours", *Journal of Mathematical Imaging and Vision* 23:67-86, 2005

[14] T. Amiaz and N. Kiryati, "Dense discontinuous optical flow via contour-based segmentation", in *Proc. ICIP 2005*, Genova, Italy, September 2005, Vol. III, pp. 1264-1267.

[15] C. L. Zitnick, N. Jojic and S. B. Kang, "Consistent segmentation for optical flow estimation", in *Proc. ICCV 2005*, Beijing, China, October 2005.

[16] F. Ranchin and F. Dibos, "Moving objects segmentation using optical flow estimation", in *Proc. Workshop on Mathematics and Image Analysis*, Paris, September 2004.

[17] R. Vidal and S. Sastry, "Segmentation of dynamic scenes from image intensities", in *Proc. IEEE Workshop on Vision and Motion Computing*, Orlando FL, December 2002, pp. 44-49.

[18] D. Cremers, "A variational framework for image segmentation combining motion estimation and shape regularization", in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Madison, Wisconsin, June 2003.

[19] A. G. Bors and I. Pitas, "Optical flow estimation and moving object segmentation based on median radial basis function network", *IEEE Trans. Image Process*., vol. 7, no. 5, pp. 693--702, May 1998.

[20] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques", *International Journal of Computer Vision* 12(1), pp. 43-77, 1994.

[21] E. Francomano, C. Lodato, S. Lopes and A. Tortorici, "An algorithm for optical flow computation based on a quasi-interpolant operator", *Journal of Computational Methods in Science and Engineering (JCMSE)* to be published.

[22] B. K. Horn and B. G. Schunck, "Determining optical flow", *Artificial Intelligence.* August 1981.

[23] M. Sezgin and B. Sankur, "Survey over image thresholding techniques and quantitative performance evaluation", *Journal of Electronic Imaging* 13(1), 146–165, January 2004.

[24] A. Abutaleb, "Automatic thresholding of gray-level pictures using two-dimensional entropy", *Computer Vision, Graphics, and Image Processing*, Volume 47, Issue 1, Pages 22-32, July 1989.

[25] M. Roggero. "Object segmentation with region growing and principal component analysis", in *Proc. ISPRS Commission III, Symposium 2002*, Graz, Austria, September 2002, pp. A-289-294.

[26] J. Fan, D. K. Y. Yau, A. K. Elmagarmid and W. G. Aref, "Automatic image segmentation by integrating color-edge extraction and seeded region growing", *IEEE Trans. Image Process.*, vol.10, no.10, pp.1454–1466, Oct. 2001.

[27] R. Adams and L. Bischof, "Seeded region growing", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.16, no.6, pp.641–647, June 1994.

[28] S. A. Hojjatoleslami and J. Kittler, "Region growing: A new approach", *IEEE Trans. Image Process.*, vol.7, no.7, pp.1079–1084, July 1998.

[29] Y. L. Chang and X. Li, "Adaptive image region-growing", *IEEE Trans. Image Process.*, vol.3, no.6, pp.868–872, Nov. 1994.