

## Erosion based Method for Quantification of Facial Palsy

M. Găianu, G. Cristescu, D. M. Onchiş

### Mihail Găianu

West University of Timișoara  
România, 300223 Timișoara, Vasile Pârvan No. 4,  
E-mail: mgiianu@info.uvt.ro

### Gabriela Cristescu

Aurel Vlaicu University of Arad  
România, 310130 Arad, Bd. Revoluției, Nr. 77,  
E-mail: gcristescu@inext.ro

### Darian M. Onchiş

University of Vienna  
Austria, Nordbergstrasse 15, A-1090, Vienna  
E-mail:darian.onchis@univie.ac.at

#### Abstract:

This paper presents a novel 3D face recognition method developed by means of erosion of a contractible topological space. The procedure is involved in the quantification of facial palsy using pre-marked points. The recognition is done in two steps: in the first step a detection of the pre-marked mimic points on a human face is performed and in the second step the mutual distances between points are calculated. A procedure of deducing the mobility or the immobility of points and the amplitude of mimic points movement follows.

**Keywords:** distance, deletable point, interest point of mimic, Univalued Segment Assimilating Nucleus.

## 1 Introduction and main results

Face recognition is a challenging task, that has been extensively investigated during the last years by many authors (e.g. [1, 2]). While most of the previous face recognition works based on 2D images, the development of 3D scanning techniques moved the domain towards the 3D face recognition. In a controlled environment, the actual 2D face recognition techniques can achieve acceptable accuracy, but inadequate for more challenging applications, especially in the presence of variation of lighting conditions and positions. Therefore, the approach using 3D recognition, by including the complete geometrical information, provides potential to alleviate the impact of lighting and position and may improve the recognition performance.

Detecting and tracking markers on the 3D human face is a challenging task. A standard procedure in this field is to use the level set algorithm ([1]). An important aspect is to measure the geodesic distances between markers applied on the 3D human face in order to emphasize the interest points of mimic and to evaluate the improvement of the face mobility. The purpose of this paper is to elaborate a method of detecting the interest points of face mimic, followed by computing the mutual distances between them together with a method of deducing the movement and its amplitude. Nowadays physicians use to measure these distances by means of the classical rule, which is an operation affected by considerable error. A computer aided procedure of points detecting and distance computing benefits of the possibility of increasing the resolution, which diminishes the error size.

The patient is placed in a two mirror system, which was developed by Frey, Jenny, Giovanoli and Stussi in ([2]). In the first step, the interest points of mimic are marked with a water-resistant pen on the patient's face (see Figure 1), obtaining a mask. The mimic interest points are:

- Central points: Central Nose (CN), Philtrum (PH=CM);
- Left side points: Left Ala of the Nose (LAN), Left Brow (LBP), Left Lower Eyelid (LLE), Left Mouth Corner (LMC), Left Midlateral Point of the Lower Lip (LML), Left Midlateral Point of the Upper Lip (LMU), Left Tragus (LTR), Left Upper Eyelid (LUE);
- Right side points: Right Ala of the Nose (RAN), Right Brow (RBP), Right Lower Eyelid (RLE), Right Mouth Corner (RMC), Right Midlateral Point of the Lower Lip (RML), Right Midlateral Point of the Upper Lip (RMU), Right Tragus (RTR), Right Upper Eyelid (RUE).

In a computer image each interest point of the mimic, i.e. each marked point of the mask, is represented by an area consisting in a set of pixels colored by the marker called Univalve Segment Assimilating Nucleus (usually denoted by USAN) [14]. Marking is necessary, because otherwise the points cannot be identified at the same place, especially on the side views of the face of the two mirror images.

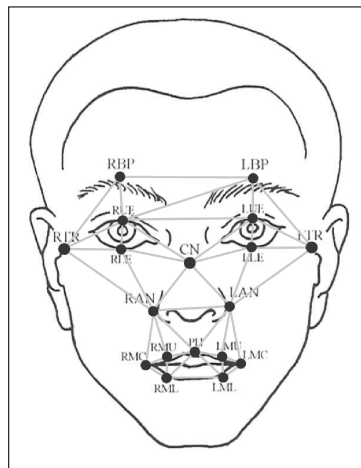


Figure 1: Pattern of the markers

Phases of diagnose computer aided process presented in this paper are:

- Operation 1: Marking interest points of mimic on patient's face;
- Operation 2: Acquiring the image of patient's face from video camera;
- Operation 3: Selecting a restricted area including one interest point of mimic;
- Operation 4: Detecting the contour of USAN (Univalve Segment Assimilating Nucleus);
- Operation 5: Determining a representing pixel of USAN by erosion and detecting a marker with the algorithm 1;
- Operation 6: Preserving the coordinates of the representing pixel;

- Operation 7: Decision: If all interest points are detected:
  - No - Repeat from Operation 3 to Operation 6;
  - Yes - Go to Operation 8;
- Operation 8: Computing mutual distances between pixels representing interest points of mimic and introducing them into the database;
- Operation 9: Repeating Operations 2 - 7 and computing the maximum variation of each distance;
- Operation 10: Decision on the mobility by comparing the variation of distances;
- Operation 11: Presenting the result to the physician.

The paper is structured as follows. In the second section the algorithm used during the proposed procedure of the computational diagnose aid is described. In the third section, after the detection (see Figure 2), we calculate the mutual distances between the markers on the face and use them to describe the face mobility. Some numerical results are included.

## 2 Active contours models and interest points detection

There are two main approaches in active contours models based on the mathematic implementation: snakes and level sets. Snakes explicitly move predefined snake points based on an energy minimization scheme, while level set approaches move contours implicitly as a particular level of a function.

As image segmentation methods, there are two kinds of active contour models according to the force evolving the contours: edge- and region-based. Edge-based active contours use an edge detector, frequently based on the image gradient, to find the boundaries of sub-regions and to attract the contours to the detected boundaries.

The first model of active contour was proposed by Kass, Witkin and Terzopoulos in ([5]) and named snakes due to the appearance of contour evolution. This method originates in various erasing techniques, which started with the algorithms of Rosenfeld ([8]). The same class of chain techniques are used in linear programming, for example in fast solving a class of transportation problems ([4]). The snake method consists in selecting a set of points, called snake points, which is connected and defines the contour. The snakes points are initially placed at further distance from the boundary of the object. Then, each point moves towards the optimum coordinates, according to some specific rule. The snake points eventually stop on the boundary of the object. The classic snakes provide an accurate location of the edges only if the initial contour is given sufficiently closed to the edges because they make use only of the local information along the contour. Estimating a proper position of initial contours without prior knowledge is a difficult problem. Also, classic snakes cannot detect more than one boundary simultaneously because the snakes maintain the same topology during the evolution. That is, snakes cannot split into multiple boundaries or merge from multiple initial contours.

For that reason more authors discuss approach using the level set method, which was first proposed by Osher and Sethian [7], [13]. A formulation to implement active contours in this way, was proposed by Osher and Sethian ([11], [7]). Level sets represent a contour implicitly via a two-dimensional Lipschitz-continuous function  $\phi(x, y) : \Omega \rightarrow \mathbb{R}$  defined on the image plane  $\Omega$ . This function results by solving a differential equation. This equation is digitized into a finite

differences equation, which describes the frontier of the domain by level sets.

The model we use in this paper takes into account the pixel by pixel description of the image, originating in the fundamental image analysis literature ([8]). Each image is represented as a  $n \times m$  matrix, where  $n$  is the number of pixels on horizontal and  $m$  is the number of pixels on the vertical of the screen. So, the image is a subset of a network  $\mathbb{Z}^2(h) = \{(ih, jh) | i, j \in \mathbb{Z}\}$  over a rectangle from  $\mathbb{R}^2$ , of sides equal to the dimensions of the computer screen. The number  $h$  is given by the selected resolution, representing the length of the side of a pixel (without loss of generality, a pixel is considered to be a square). So, the structure in which we work is that of a totally bounded metric space embedded into the linear space  $\mathbb{R}^2$ . Also, we consider it as a topological space, with the topology induced by the metric. The number  $h$  is considered to be the measurement unit for computing and decision making and it is finally converted into the desired unit according to the requirement of the user.

A subset of an image, which is the object of our computation procedure, is the digitization of a contractible topological space from mathematical point of view, i.e. a topological space that reduces to a point by means of a homotopy. The following concepts and properties are necessary in order to describe the procedures. Let  $X \subseteq \mathbb{R}^2$  a contractible topological space. We consider, without loss of generality, the digitization of the space  $X$  by means of the method of digitization,  $f: \mathbb{R}^2 \rightarrow \mathbb{Z}^2$ ,

$$f(x, y) = ([x + \frac{1}{2}], [y + \frac{1}{2}]), (x, y) \in \mathbb{R}^2.$$

If  $x(ih, jh) \in \mathbb{Z}^2(h)$  is a grid point then its 4-neighborhood is the set  $V_4(x) = \{x, (ih, (j-1)h), (ih, (j+1)h), ((i-1)h, jh), ((i+1)h, jh)\}$  and its 8-neighborhood is the set  $V_8(x) = V_4(x) \cup \{((i-1)h, (j-1)h), ((i+1)h, (j-1)h), ((i-1)h, (j+1)h), ((i+1)h, (j+1)h)\}$ . The concepts of connectivity used in this paper are the 4-connectivity (i.e. defined using  $V_4(x)$  by horizontal and vertical lines) and the 8-connectivity (i.e. defined using  $V_8(x)$  by horizontal, vertical, at  $45^\circ$  and at  $135^\circ$  lines) by means of arcs, as defined in [8].

**Proposition 1.** *The digital image  $f(X)$  of a contractible topological space  $X \subseteq \mathbb{R}^2$  is a 8-connected set.*

**Proof.** If  $X$  is contractible then it is path connected. Denote by  $S = f(X) \subset \mathbb{Z}^2(h)$  and consider the 8-connectivity in  $S$ . Suppose that  $S$  has an isolated point  $a = (kh, lh)$ ,  $k \in \mathbb{N}$ ,  $l \in \mathbb{N}$ . Let us consider the pixel having the center  $(kh, lh) = f(a)$ , i.e. the square  $D = [kh - h/2, kh + h/2) \times [lh - h/2, lh + h/2)$ . There is at least a point from  $X$  situated in the rectangle  $D$ , say point  $b$ . Therefore  $f(b) = a$ . Since  $a$  is an isolated point of  $S$ , The squares of side  $h$  and centers  $((k-1)h, (l-1)h)$ ,  $((k-1)h, lh)$ ,  $((k-1)h, (l+1)h)$ ,  $(kh, (l-1)h)$ ,  $(kh, (l+1)h)$ ,  $((k+1)h, (l-1)h)$ ,  $((k+1)h, lh)$ ,  $((k+1)h, (l+1)h)$ , i.e. the 8-neighbors of  $a$ , are not in  $S = f(X)$ . So, none of these squares contain points of  $X$ . Therefore,  $X$  is not connected, which contradicts the hypothesis that space  $X$  is contractible.  $\diamond$

Let  $S \subset \mathbb{Z}^2$  a finite subset. After removing a point  $x = (i, j) \in S$  from  $S$  the number of connected components of  $S$  or of  $\mathbb{Z}^2 \setminus S$  may increase or decrease. Let us suppose that  $S$  is 4-connected (the case of the 8-connectivity is similar). A. Rosenfeld proved in [8] that if  $S \subset \mathbb{Z}^2$  is a finite subset,  $S \setminus \{x\}$  does not have less 4-connected components than  $S$  if and only if  $V_4(x) \cap S \neq \emptyset$ . Also,  $\mathbb{Z}^2 \setminus (S \setminus \{x\})$  has not more 8-connected components than  $\mathbb{Z}^2 \setminus S$  if and only if  $V_8(x) \cap (\mathbb{Z}^2 \setminus S) \neq \emptyset$ . If  $x \in S$  such that  $V_4(x) \cap S$  is contained in a 4-connected component of  $V_8(x) \cap S$ , then  $S \setminus \{x\}$  does not have more 4-connected components than  $S$ . More, if  $\mathbb{Z}^2 \setminus S$  does not have more 4-connected components than  $S$ , then  $V_4(x) \cap S$  is included in a 4-connected component of  $V_8(x) \cap S$ . As consequence, the following concept is introduced in [8].

**Definition 2.** An element  $x \in S$ , which verifies the conditions: 1)  $V_4(x) \cap S \neq \emptyset$ , 2)  $V_8(x) \cap (\mathbb{Z}^2 \setminus S) \neq \emptyset$ , and 3)  $V_4(x) \cap S$  is included in a 4-connected component of  $V_8(x) \cap S$ , is called a 4-deletable element.

As consequence of the above mentioned properties, Rosenfeld has proven in [8] that:

**Theorem 3.** Let  $S \subset \mathbb{Z}^2$  a finite set with the property that  $\mathbb{Z}^2 \setminus S$  has only one 8-connected component. Then  $S \setminus \{x\}$  has the same number of 4-connected components as  $S$  and  $\mathbb{Z}^2 \setminus (S \setminus \{x\})$  has an unique 8-connected component if and only if  $x$  is 4-deletable.

In [8] the deletable elements of a set are classified into two classes: extremities and corners. A point  $x \in S$  is called extremity if  $\text{card}(V_4(x) \cap S) = 1$ . A point  $x \in S$  is said to be a corner if  $V_4(x) \cap S$  has exactly two neighbors of  $x$  in  $S$ . Every simple 4-connected set  $S$  having more than one element has at least two deletable elements.

In case of the 8-connectivity, a point  $x \in S$  is said to be an extremity if  $V_8(x) \cap S$  has either exactly one element, or two 4-neighboring elements, or three 4-neighboring elements out of which only one is in  $V_4(x)$ . A point  $x \in S$  is said to be a corner if  $V_8(x) \cap S$  has exactly three, four or five 4-neighboring elements, out of which only two are in  $V_4(x)$ .

The Diagnose Procedure consists in a succession of computer programs, as it follows:

- i. *Program for frames acquiring* divides the succession of pictures taken from the face of the patient into frames, called images. The film is of AVI (Audio Video Interleaved) format, divided into frames having PNG (Portable Network Graphics) format, in order to allow procession. The resolution of the film is  $720 \times 576$  pixels, with  $25fps$  (frames per second). The frames are coded either on 8 bits (PNG-8) or on 24 bits (PNG-24), including the transparency on 256 levels. The result is a collection of 4150 images, indexed by  $i \in \{1, 2, \dots, 4150\}$ .
- ii. *Program for detection of the mimic interest centers on patient's face, and computation of mutual distances between them.* The pixel by pixel erosion algorithm below presents the manner of processing each image in order to detect the coordinates of each central pixel representing each marker on patient's face. Also, the manner of validation and storage of each marker is included. The search of these points is optimized in the algorithm we cut the margins, which do not offer relevant information. In this manner we reduce the number of color tests. Each center of marker is the center of an USAN area, which is approximately ellipse (circle) shaped. The USAN area is detected based on pixel by pixel color tests, detecting the contour and erasing each USAN area following the contour. The erosion means changing the color from the artificial USAN area color into face color. After determining the coordinates of centers of all USAN areas from an image, the mutual distances between these pixels is computed and stored into a database. The database contains all these distances taken from all the 4150 images from our collection.
- iii. *Program for computing the difference between an image and the initial one.* Initially the patient's face is in relaxed position, which gives the so called initial image. In order to take the other images from the collection, the patient is asked to perform various movements of the face. The differences between the corresponding distances between the mimic interest centers give the information on the existence of a displacement on the face from an image to another one. The maximum positive difference and the minimum negative difference are chosen from the database, together with the images in which they appear. These images are presented to the physician, together with the numerical results, converted into required units.

- iv. *Decision on the movement variability* is taken by analyzing the mutual distances and comparing them from an image to the next one. If the distance from a center, for example CN to another one, for example RAN does not change and the rest of the distances from LAN to LTR, CN, RAN are variable, then it follows that LAN is the moving point. It means that the patient moved ones face. The decision is also stored for further investigations and comparisons in order to assess the evolution of patient's face mobility in various stages of the treatment.

---

**Algorithm 1: Detecting a point on the patient's face**


---

```

Data: imagine, contour, x, y, area, detect
Result: detection of the markers on the patient face
1 initialization;
2 eps=2;
3 ok=0;
4 cnt;
5 while ok<25 do
6   ok=0;
7   cnt=contor;
8   for j=x-20;j<=x+20;j++ do
9     for i=y-30;i<=y+30;i++ do
10      pixel=(rgb*)(data+i*width+j);
11      is[cnt]=i;
12      js[cnt]=j;
13      vals[cnt]=area;
14      if valid(imagine,pixel,i,j,eps)==1 then
15        vals[cnt]=detect;
16        ok++;
17      end
18      cnt++;
19    end
20  end
21  eps+=3;
22 end
23 printf("%d\n",eps);
24 contor=cnt;

```

The algorithm 1 detects the markers LTR, RTR, CN and others. The marker is being searched in a area of  $40 \times 60$  pixels.

### 3 Numerical results

As a result of Operation 6 one obtains a collection of distinct points on the face of the patient.

Usually physicians measure the distance between each pair of these points using the rule, i.e. using the Euclidean distance in two dimensions. This is the reason we use this distance in our experiment. Increased accuracy may occur if another kind of metric is used, for example the Earth Mover's Distance. The Earth Mover's Distance was introduced in the computer vision

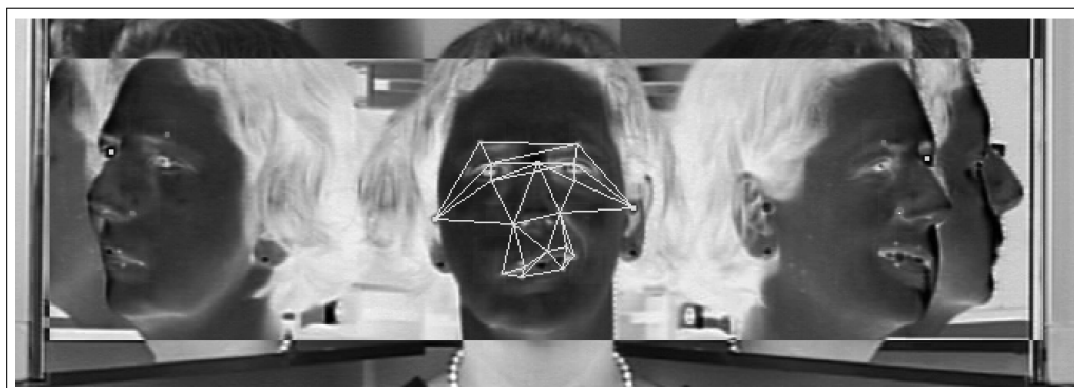


Figure 2: Detection of the points on the patient's face

Table 1: Mutual distances between the interest points of face mimic from figure 2

Location of interest points	Mutual distance
RTR↔RBP	55.569776
RTR↔RUE	49.497475
RTR↔RLE	42.438190
RTR↔RAN	49.091751
LTR↔LBP	53.150729
LTR↔LUE	46.400431
LTR↔LLE	40.718546
LTR↔LAN	46.097722
CN↔RUE	29.000000
CN↔RLE	31.016125
CN↔RAN	40.853396
CN↔LUE	23.000000
CN↔LLE	25.495098
CN↔LAN	34.014703
RBP↔LBP	61.008196
RBP↔RUE	14.764823
RUE↔LBP	55.317267
RUE↔RLE	11.000000
RLE↔RAN	30.413813
RLE↔LUE	53.150729
LBP↔LUE	12.165525

community in 1997 by Rubner, Guibas and Tomasi [10], [9] to overcome inconsistencies with perceptual similarity observed in (weighted)  $L_p$  -norms and quadratic forms.

The innovation of the paper consists in a more elaborated face model which allows quantification of facial palsy. Based on this model, a new scoring system for mimic function recovery is developed. More accurate results which are relevant for plastic surgery as well as for face models in computer vision are obtained. Another advantage of the method described in this paper is the real time data acquisition and processing. A fast high resolution real time 3D video system is used to acquire data from patients. They are filmed and analyzed before and after any kind of treatments in a standardized pre-treatment and follow-up scheme. If a patient is repeatedly analyzed in the above described manner, a database is constructed, which exactly document and visualize the recovery course of facial palsy. The system is able to recognize shape, texture, movements and illumination conditions of the face and is automatic in acquiring and processing patient's data. This simplifies the process of measuring and - at the same time - enhance the clinical methodology in terms of diagnostic and prognostic accuracy and convenience for patients. For example, the mutual distances between the mimic interest points on the face from figure 2 are in table 1. They are presented to the physician to be used in medical diagnose process. In this case the color to detect, i.e. the markers' color, is green and the function to maximize is a three dimensional vectorial one, obtained based on a normalized RGB description, in which the first coordinate refers to green. The order relation in  $\mathbb{R}^3$  is the lexicographic one.

An important consequence of the technique described in this paper is the automatic evaluation of the facial palsy's degree of the patient. The system is able to directly perform a 3D face recognition thus supporting medical staff in exact planning the patient's medical procedures. The models allow to determine the position and movements of facial features with high accuracy, avoiding model bias in fitting to particularities of individual faces and diseases. The procedure described in this paper is setting up a high-end up to date standardized 3D facial palsy grading system, which is been used to monitor and document post-treatments course for facial palsy patients and to compare various surgical outcomes from different centers around the world. In fact, we propose a software that may lead to performing a standardized grading system and a comparison of 3D human face data.

The detection and tracking techniques may be extended from image data to video data. Since video is more complex, even more efficient retrieval techniques need to be developed using a multi-projection system as in [6], applying the powerful perception-oriented Earth Mover's Distance to video data.

## 4 Conclusions and Future Works

Detection in large image or other multimedia databases highly depends on the underlying similarity model. In future work, we plan to extend our detection and tracking concept techniques from image data to video data. Since video is more complex, even more efficient retrieval techniques need to be developed using a multi-projection system as in ([6]).

We plan to employ generative face models (i.e., models that can produce synthetic face images) which can be used for visualization of prognosis appearance of the patient's mimic and facial movements. Predictions of facial movements can be based on the statistical analysis of empirical data such as the past progression (e.g., preoperative/postoperative appearance) of facial movements of the subject or the class of subjects sharing similar pathologies. The ability to generate synthetic images of a face can further be exploited for biofeedback training of mimic musculature. Another important aspect in the future work will be the ability of quantifying and assessing mimic function by spontaneous facial movements as well.



## Bibliography

- [1] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino, 2d and 3d face recognition: A survey, *Pattern Recognition Letters*, 28(14):1885 - 1906, 2007.
- [2] M. Frey, A. Jenny, P. Giovanoli, and E. Stussi, Development of a new documentation system for facial movements as a basis for the international registry for neuromuscular reconstruction in the face, *Plast Reconstr Surg*, 93(7):1334-1349, 1994.
- [3] M. Găianu, *Contributions to the research of some problems of image recognition and retrieval (in Romanian)*, Thesis, West University of Timișoara, 2011.
- [4] P.L. Hammer, S. Rudeanu, *Boolean methods in operations research and related areas*, Springer-Verlag, 1968.
- [5] M. Kass, A. Witkin, and D. Terzopoulos, Snakes, active contour model, *International Journal of Computer Vision*, 321-331, 1988.
- [6] D. Onchis and H. Feichtinger. Constructive reconstruction from irregular sampling in multi-window spline-type spaces, *General Proceedings of the 7th ISAAC Congress*, London, 2010.
- [7] S. Osher and J. A. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulations, *Journal of Computational Physics*, 79(1):12 - 49, 1988.
- [8] A. Rosenfeld, Connectivity in Digital Pictures, *J. ACM*, 17(1):146-160, 1970.
- [9] Rubner, Y., Guibas, L. and Tomasi, C. The Earth Mover's Distance, Multi-Dimensional Scaling, and color based image retrieval. In *Proc. of the ARPA Image Understanding Workshop*, 661-668, 1997.
- [10] Rubner, Y. and Tomasi, C., *Perceptual Metrics for Image Database Navigation*, Kluwer Academic Publishers, 2001.
- [11] J. A. Sethian, A fast marching level set method for monotonically advancing fronts, *Proceedings of the National Academy of Sciences of the United States of America*, 93(4):1591-1595, 1996.
- [12] J. A. Sethian. *Level set methods and fast marching methods - evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science*. In *Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science*. Cambridge University Press, 1998.
- [13] J. Sethian, *Level Set Methods and Fast Marching Methods*. Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press, 2nd ed., 1999.
- [14] S.M. Smith, J.M. Brady, SUSAN - a new approach to low level image processing, *Int. Journal of Computer Vision*, 23(1):45-78, 1997.