CVPR-2000 international conference on Computer Vision and Pattern Recognition - pp. 685-690 - Hilton Head Island, USA - 13-15 June, 2000

Robust Visual Recognition of Colour Images

Rozenn Dahyot
LRS
11 rue Jean Mentelin
Strasbourg, 67035 FRANCE
Rozenn.Dahyot@equipement.gouv.fr

Pierre Charbonnier
LRS
11 rue Jean Mentelin
Strasbourg, 67035 FRANCE
Pierre.Charbonnier@equipement.gouv.fr

Fabrice Heitz
LSIIT UPRESA CNRS 7005
Université Strasbourg 1
Illkirch, 67400 FRANCE
fabrice.heitz@ensps.u-strasbg.fr

Abstract

In this paper, a robust pattern recognition system, using an appearance-based representation of colour images is described. Standard appearance-based approaches are not robust to outliers, occlusions or segmentation errors. The approach proposed here relies on robust M-estimators, involving non-quadratic and possibly non-convex energy functions. To deal with the minimisation of non-convex functions in a deterministic framework, we introduce an estimation scheme relying on M-estimators used in continuation, from convex functions to hard redescending non-convex estimators. At each step of the robust estimation scheme, the non-quadratic criterion is minimized using the half-quadratic theory. This leads to a weighted least squares algorithm, which is easy to implement. The proposed robust estimation scheme does not require any user interaction because all necessary parameters are previously estimated. The method is illustrated on a road sign recognition application. Experiments show significant improvements with respect to standard estimation schemes.

1 Introduction

Appearance-based representation of objects has recently received considerable attention [10, 9, 20, 4, 16, 14, 17]. One popular approach is the eigenspace representation [10, 9, 20, 4, 11] which allows a substantial dimensionality reduction of the recognition problem. Eigenspace methods involve a reconstruction procedure, which consists in projecting the observation on the training eigenspace. Then, this projection is identified with the closest model of the database. Traditional least-squares (LS) estimation which corresponds to the orthogonal projection on the eigenspace, is sensitive to gross errors (outliers) that occur, for instance, when the object is partially occluded [10].

To deal with outliers and occlusions, recent studies have proposed local appearance-based representations of objects [11, 20, 4]. Instead of learning the whole object appearance, small parts are used as training images. Robustness of those local appearance-based matching methods is based on the assumption that at least one sub-region chosen for the LS-estimation, is not corrupted by outliers [20, 4, 11]. One drawback of local eigenspace-based representation is that more parameters are required to reconstruct (or index) the whole image.

Another way to cope with outliers, is to reformulate the reconstruction step as a robust estimation problem [2, 7]. Among the methods proposed in the context of robust statistics [6, 19, 18, 8, 15], M-estimators offer a good compromise between algorithmic complexity and outlier rejection capability. Their use for eigenspace recognition was first introduced by Black [2]. Robust estimation can be applied either with local or global eigenspace representation of appearance.

In this work, we use a global appearance representation of objects [10]. A robust estimation is performed on the training eigenspace using M-estimators [6, 2]. The first contribution of this paper is to formulate M-estimation in the framework of the half-quadratic theory [5, 3](section 2). This theory introduces an auxiliary variable which, in our case, can be interpreted as an outlier mask. This provides a natural linearization of the normal equations, and results in a - fast and easy to implement - weighted least squares algorithm.

Parameter estimation is an important point for image reconstruction but is seldom addressed in the literature. Most of the time, the tuning of parameters is let to the user. As a second contribution of this paper, we propose a method to estimate the scale parameter of the M-estimators (section 4), making the whole method data driven. We also extend the method to colour image recognition, taking colour components into account in a pixel-by-pixel fashion. This is explained in section 3.

Finally, we apply our method to hard redescending non-convex M-estimators, which provide better outlier rejection. Following the same idea as in the Graduated Non Convexity (GNC) [2] algorithm, we gradually introduce non-convexity by using three M-estimators in continuation. This results in a new, simple to implement and non-supervised robust recognition scheme for colour images which is applied to road sign recognition (section 6).

2 Robust estimation using the half-quadratic theory

To simplify, we first consider the case of grey level images. The extension to RGB images is presented in section 3. The training images are arranged as n-dimensional vectors by lexicographic ordering. A Principal Component Analysis (PCA) is performed and t eigenvectors are retained to span the eigenspace F ($t \ll n$). Given an unknown image e, standard eigenspace techniques perform its reconstruction on F by computing the best representative e^* of e as a linear combination of eigenvectors :

$$e^* = \sum_{j=1}^{t} c_j U_j \tag{1}$$

where c_j is the j^{th} unknown co-ordinate of e^* on the j^{th} eigenvector U_j. For grey level images, the residual on the i^{th} pixel is defined by:

$$\epsilon_i = e_i - e_i^* \tag{2}$$

Least Squares estimation. A standard method for the estimation of e^* consists in minimizing the quadratic norm $J_0 = \|\epsilon\|$. Geometrically, the solution $\hat{e^*}$ of the least squares estimation of e, is the orthogonal projection of e onto the t-dimension subspace F. As it is well known, least squares estimation is sensitive to gross errors (outliers) produced, for instance, by occlusions [2].

M-estimation in the Half-Quadratic framework. M-estimators, involving non-quadratic and possibly non-convex energy functions, are naturally robust to outliers or gross errors. M-estimation leads to the minimisation of a robust norm J_1:

$$J_1(c) = \sum_{i=1}^{n} \rho(\epsilon_i) \tag{3}$$

where function ρ increases at a lower rate than the quadratic function. To minimise J_1 in a deterministic framework, we propose to use the *Half-Quadratic* theory [5, 3]. Under certain conditions on ρ [3], the non-quadratic energy J_1 is transformed into an augmented energy by introducing an auxiliary variable b.

$$\min_c \left\{ J_1(c) = \sum_{i=1}^{n} \rho(\epsilon_i) \right\} = \min_c \min_b \left\{ J_1^\sharp(c,b) = \sum_{i=1}^{n} (b_i . \epsilon_i^2 + \beta(b_i)) \right\} \tag{4}$$

where β is a function of b_i. $J_1^\sharp$ is *half-quadratic*, i.e.:

- When b is fixed, $J_1^\sharp$ reduces to a weighted least-squares criterion, whose solution satisfies:

$$(U^T \cdot B \cdot U)c = U^T \cdot B \cdot e \tag{5}$$

  where $B = diag_{i=1\dots n}\{b_i\}$. There are many possible numerical algorithms to solve (5) (the conjugate gradient algorithm [13] is used here).

- When c is fixed, $J_1^\sharp$ becomes *convex* with respect to b. Moreover, it can be shown [3] that the explicit minimizer is given by $b_i = b(\epsilon_i) = \frac{\rho'(\epsilon_i)}{2 \cdot \epsilon_i} \quad \forall i$. Due to the properties of ρ [3], b_i is close to one when ϵ_i is small (inliers), and vanishes for large values of ϵ_i (outliers). Therefore, b can be seen as an *outlier mask*, conceptually similar to the one defined in [2], excepted that:
  - the mask is not Boolean in our case : b_i is a real between 0 and 1,
  - b appears naturally in our formulation and participates to the minimization process.

Given an initial guess c_0, we use the following alternate minimization algorithm:

$$\left\{ \begin{array}{l} \forall i \in \{1 \dots n\} \; \epsilon_i^{(m)} = e_i - \sum_{j=1}^{t} c_j^{(m)} \cdot U_{ij} \\ \forall i \in \{1 \dots n\} \; b_i^{(m+1)} = b(\epsilon_i^{(m)}) \\ (U^T \cdot B^{(m+1)} \cdot U) \cdot c^{(m+1)} = U^T \cdot B^{(m+1)} \cdot e \end{array} \right. \tag{6}$$

We can notice that algorithm (6) is similar to the Location Step With Modified Weights, proposed by Huber in the context of robust statistics ([6] p.183). The half-quadratic theory is valid for a large class of functions, defined in [3]. Three standard robust functions, considered in this paper, are presented in Table 1.

	$\rho(x)$	$\rho'(x)$	convexity
HS	$2\sqrt{1+x^2}-2$	monotone	convex
HL	$\log(1+x^2)$	soft redescender	non convex
GM	$\frac{x^2}{1+x^2}$	hard redescender	non convex

Table 1. ρ **functions**

3 Extension to colour image

We now consider the case of colour images. Colour images are transformed into 1-D vectors by concatenating red, green and blue values and a PCA is applied to the $3n \times 3n$ covariance matrix. A first important point for the reconstruction is the definition of the residual ϵ_i. Since colour is a discriminant feature for recognition, the red r_i, green g_i and blue b_i components of the pixel i must not be considered separately. We consider colour components in a pixel-oriented fashion and define the residuals as:

$$\epsilon_i = [(r_i - r_i^*)^2 + (g_i - g_i^*)^2 + (b_i - b_i^*)^2]^{\frac{1}{2}} \qquad (7)$$

It is possible, as before, to take into account outliers, using the theory previously described in section 2, and the resulting algorithm is similar to (6):

$$\begin{cases} \forall i \in \{1 \dots n\} \; \epsilon_i^{(m)} \; is \; calculated \; using \; (7) \\ \forall i \in \{1 \dots n\} \; b_i^{(m+1)} = b_{i+n}^{(m+1)} = b_{i+2n}^{(m+1)} = b(\epsilon_i) \\ (U^T \cdot B^{(m+1)} \cdot U) \cdot c^{(m+1)} = U^T \cdot B^{(m+1)} \cdot e \end{cases} \qquad (8)$$

4 Estimation of the scale parameter

The robust estimator defined by equation (3) usually depends on a *scale parameter* σ_ρ, that controls the point where the influence of outliers begins to decrease:

$$J_1(c, \sigma_\rho) = \sum_{i=1}^{n} \rho \left(\frac{\epsilon_i}{\sigma_\rho} \right) \qquad (9)$$

Several methods have been proposed to adjust the scale parameter σ_ρ [6, 19]: joint estimation of (c, σ_ρ) [6] or estimation of σ_ρ before reconstructing c. A joint estimation of (c, σ_ρ) is generally computationaly demanding. For computational efficiency, we first estimate the scale parameter *offline*, using the images of the database. We then perform the reconstruction step, keeping the scale parameter fixed. This approach has revealed both fast and robust. Assuming that the learning database is representative, on a statistical point of view, for each pixel i, the expected error ϵ_i between the observation and the reconstruction is only due to the truncation on the t-dimensional eigenspace. In practice, we compute the variance of the residuals for each image k in the database: $\sigma_{k,t}^2 = \frac{1}{n}\sum_{i=1}^{n}(\epsilon_i^{k,t})^2$ where t designates the number of eigenvectors used to describe the eigenspace and $\epsilon_i^{k,t}$ is defined as in (2) (for grey level images) or as in (7) (for colour images). In practice, the distribution of residuals is well approximated by a gaussian density. As a consequence, about 95% of the computed residuals belong to the interval $[0; 2\sigma_{k,t}]$. Considering the whole database (composed of N training images), it comes that more than 95% of the residuals are contained in $[0; 2\sigma_t]$ where:

$$\sigma_t = \max_{k \in [1 \cdots N]} \{\sigma_{k,t}\} \qquad (10)$$

Considering that points i with residuals $|\epsilon_i^t| > 2\sigma_t$ are outliers and that for non-convex robust functions ρ, the influence of the outliers begins to decrease at their inflexion point [2], one derives the following scale parameters : $\sigma_{\rho_{HL}} = 2\sigma_t$ for the non convex HL function (see Table 1), and $\sigma_{\rho_{GM}} = 2\sqrt{3}\sigma_t$ for the GM function. For the convex function HS, we have fixed arbitrarily the scale parameter to $\sigma_{\rho_{HS}} = 2\sigma_t$, by considering that the behaviour of HS changes from quadratic to linear at point $x = 1$. Once σ_ρ is computed, the minimization can be performed using algorithm (8) and modifying the weights according to $b_i = b(\frac{\epsilon_i}{\sigma_\rho})$.

5 M-estimation in continuation

Convex functions, like HS, yield a unique solution, but the corresponding *influence functions* ρ' are monotone. The influence of outliers is thus bounded, but not null. From this point of view, hard redescenders [19], such as GM, are much more attractive. Unfortunately, hard redescenders yield highly non-convex objective functions. Efficient deterministic algorithms can, however, be defined in this case, using a gradual approach of non-convexity. We propose here to use (in continuation) the convex function HS, (using the least square estimate c_{LS} as an initial guess) followed by the non-convex soft redescender HL, and by the the non-convex hard redescender GM. This strategy is similar to GNC (used in [1, 2] for object recognition, with the GM function). However, our approach shows a major improvement: by exploiting the half-quadratic theory, the algorithm explicitly addresses the problem of the non-quadraticity of the energy function. This approach yields a natural linearization of the normal equations.

6 Experimental Results

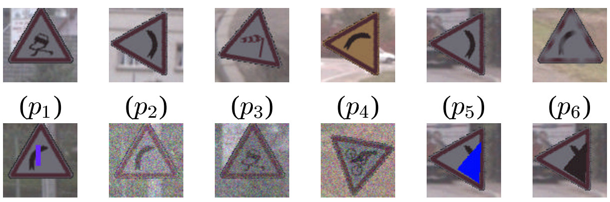
To assess the performances of the proposed approach, we have tested our method on a road sign recognition problem. The set of (european) triangular road signs has been learned, along with their rotation in the image plane. We consider that the recognition is good when the correct road sign is identified, with the correct rotation. The robust estimators in continuation have been tested against several artefacts like cluttered backgrounds, occlusions and white noise.

Training database. A training set of 1548 (76×76) colour images representing 43 different road signs has been collected (36 images per road sign, one every 10 degrees in the image plane). A selection of training images is represented in figure 1. Only 60 eigenvectors are kept in this experiment to span the eigenspace, which represents about 90% of the information in the database [10]. After each estimation (standard LS and M-estimations with functions HS, HL and GM), the euclidian distance between the estimation e^* and all the training images in the eigenspace is computed and the closest model is selected.

t_1 t_2 t_3 t_4 t_5 t_6 t_7

Figure 1. Several road signs

Test images. The robust recognition method has been tested on images with large corrupted areas (figure 2): cluttered backgrounds (images p_1 to p_5), cluttered background with occlusions (p_6, p_7, p_{11} and p_{12}), and cluttered backgrounds with white noise (p_8 to p_{10}). Because of the triangular shape of the road signs, the outlying cluttered backgrounds represent about 50% of the area of the test images. Images p_8 to p_{10} are corrupted by additive white gaussian noise (with SNRs of $-3.11dB$, $2.83dB$ and $1.88dB$ respectively). p_9 is the same as p_1 with noise. p_{11} and p_{12} correspond to p_5, with a large occlusion area (blue area on p_{11}, black area on p_{12}).



(p_1) (p_2) (p_3) (p_4) (p_5) (p_6)
(p_7) (p_8) (p_9) (p_{10}) (p_{11}) (p_{12})

Figure 2. Some test images.

Two parameters are defined to assess the quality of the recognition. Parameter d_1 is defined as the euclidian distance between the estimate e^* and the closest training image in the eigenspace. Designating by d_2 the distance between e^* and the second closest model in the database, we define a recognition contrast C as $C = 100 \times (1 - \frac{d_1}{d_2})$. When $C = 0$, then $d_1 = d_2$ and the identification of the observation is not reliable. When $C = 100$, then $d_1 = 0$, and the estimation e^* corresponds exactly to a model of the database.

Recognition results. For all tested images, the standard LS reconstruction and recognition approach gives wrong solutions. The values of d_1 and C are reported in Table 2 and wrong recognitions are marked as "w". Excepted for p_6 and p_8, d_1 decreases and C increases, after each step of the robust estimation procedure .

	HS	HL	GM
p_1	$d_1 = 4\,782\,740$ $C = 17$	$d_1 = 137\,516$ $C = 87$	$d_1 = 49\,119$ $C = 95$
p_2	$d_1 = 4\,089\,292$ $C = 15$	$d_1 = 43\,461$ $C = 95$	$d_1 = 16\,469$ $C = 98$
p_3	$d_1 = 5\,424\,270$ $C = 15$	$d_1 = 117\,554$ $C = 89$	$d_1 = 45\,238$ $C = 96$
p_4	$d_1 = 4\,787\,767$ $C = 24$	$d_1 = 85\,628$ $C = 94$	$d_1 = 30\,482$ $C = 98$
p_5	$d_1 = 3\,548\,091$ $C = 17$	$d_1 = 84\,858$ $C = 91$	$d_1 = 30\,482$ $C = 98$
p_6	$d_1 = 7\,107\,054$ $C = 1$	$d_1 = 1\,128\,597$ $C = 20$	$d_1 = 1\,324\,429$ $C = 20$
p_7	$d_1 = 4\,532\,272$ $C = 12$	$d_1 = 270\,583$ $C = 80$	$d_1 = 85\,411$ $C = 94$
p_8	$d_1 = 7\,360\,926$ $C = 13$	$d_1 = 4\,550\,645$ $C = 24$	$d_1 = 10\,931\,307$ $C = 12$
p_9	$d_1 = 9\,362\,029$ $C = 8$	$d_1 = 1\,352\,827$ $C = 40$	$d_1 = 416\,755$ $C = 67$
p_{10}	$d_1 = 9\,029\,500$ $C = 2$	$d_1 = 1\,146\,891$ $C = 40$	$d_1 = 297\,365$ $C = 73$
p_{11}	$d_1 = 7\,515\,481$ $C = 0.5$ (w)	$d_1 = 453\,189$ $C = 53$	$d_1 = 354\,700$ $C = 70$
p_{12}	$d_1 = 7\,979\,003$ $C = 14$ (w)	$d_1 = 852\,356$ $C = 76$ (w)	$d_1 = 677\,491$ $C = 79$ (w)

Table 2. Distance d_1 and contrast C after each robust estimation step ("w" indicates wrong recognition).

For images p_1 to p_5, with cluttered background only, the recognition results are excellent, with a recognition contrast not less than 95 after the final GM step.

For the test images with additive white noise, the robust estimation scheme recovers the correct model but the estimates are farther off from the solution, as expected (compare for instance p_9 and p_1).

For all images with occluded road signs, excepted p_{12}, a correct recognition is achieved. Let us notice that the black occlusion area in p_{12} may either be considered as inliers for the training image t_5, or as outliers for t_2. p_{12} is indeed ambiguous, even for human perception. There is no such ambiguity in p_{11}, since the blue pixels are clearly identified as outliers, except for the initial HS estimation step, which provides a wrong recognition. The small value of C indicates that $e^*_{HS}(p_{11})$ is at equal distance between two training images (the second closest model is the correct one). This wrong estimation is corrected by the HL and GM steps, which reject more strongly outliers. This example shows that colour information is taken into account with profit in the recognition and that the continuation scheme is useful, in case of ambiguous data.

As an illustration, Figure 3 shows the reconstructions, after each estimation step (LS, HS, HL and GM), for image p_2. The last image $e^*_{LS}(t_2)$ corresponds to the LS reconstruction of the correct solution t_2, on the 60 eigenvectors basis. As can be seen, the estimates improve after each step of the robust reconstruction, and the final reconstruction $e^*_{GM}(p_2)$ is close to the best estimate $e^*_{LS}(t_2)$.

$e^*_{LS}(p_2)$ $e^*_{HS}(p_2)$ $e^*_{HL}(p_2)$ $e^*_{GM}(p_2)$ $e^*_{LS}(t_2)$

Figure 3. Reconstructions of p_2 at each step of the recognition process and the LS reconstruction for the corresponding solution.

In figure 4, we show typical outlier maps (i.e. images of $\mathbf{b}$ obtained after the final estimation with function GM). Outliers correspond to dark areas (where b_i is close to 0) and inliers to light areas (b_i is close to 1). Some pixels inside the road sign appear as outliers because the truncated eigenspace representation does not allow a perfect reconstruction of the symbols.

p_1 p_2 p_6 p_7 p_{11} p_{12}

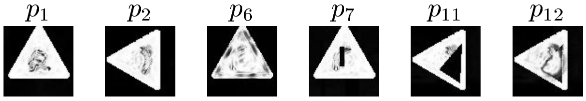


Figure 4. Some outlier maps for GM estimation.

Let us finally notice that the theoretical breakdown point[1] of M-estimators is less than $\frac{1}{1+t}$ [6, 2]. Therefore the percentage of outliers which can theoretically be handled in our case is less than 2% (with $t = 60$). However, in our experiments, we have at least 50% outliers and sometimes up to 65% corrupted data (image p_{11}). The robust method thus significantly outperforms its theoretical performances.

We have recently implemented an optimized version of the algorithm. The reconstruction time is now less than 1 second per test images on a Pentium Pro 200 MHz P.C.

7 Conclusion

We have presented a robust eigenspace recognition method for colour images, using M-estimators in continuation. Exploiting the Half-Quadratic theory, we propose a non-supervised algorithm which is simple to implement. Experiments on highly corrupted colour images of road signs, show that the method significantly outperforms its theoretical breakdown point, yielding reliable recognition in adverse situations (occlusion, noise or cluttered background).

References

[1] M.J. Black and P. Anandan, “A Framework for the robust estimation of optical flow”, Proc. of the Fourth International Conf. on Computer Vision (ICCV 93), pp. 231-236, Berlin, Germany, May 1993.

[2] M. J. Black and A. D. Jepson, “EigenTracking: Robust Matching and Tracking of Articulated Objects Using a View-Based Representation”, Int. J. Computer Vision, vol. 26, no. 1, pp. 63-84,1996.

[3] P. Charbonnier, L. Blanc-Féraud, G. Aubert and M. Barlaud, “Deterministic Edge-Preserving Regularization in Computed Imaging”, IEEE Transaction Image Processing, vol. 6, no. 2, pp. 298-311, February 1997.

[4] V. Colin de Verdiere, J.L. Crowley, ”Visual Recognition using Local Appearance”, Proc. of Fifth European Conf. on Computer Vision (ECCV 98), Freiburg, Germany, June 1998.

[5] D. Geman and C. Yang, “Nonlinear image recovery with half-quadratic regularization and FFT’s”, IEEE Trans. Image Processing, vol. 4, no. 7, pp. 932-946, July 1995.

[6] P.J. Huber, “Robust Statistics”, John Wiley & Sons, New York, 1981.

[1] the breakdown point corresponds to the percentage of outliers that estimators can tolerate before the solution becomes arbitrarily bad

[7] A. Leonardis, H. Bischof and R. Ebensberger, ”Robust recognition using eigenimages”, Technical Report PRIP-TR-47, PRIP, TU Wien, 1997.

[8] P. Meer, D. Mintz, A. Rosenfeld and D.Y. Kim, “Robust Regression Methods for Computer Vision: A Review”, International Journal of Computer Vision, vol. 6, no. 1, pp. 59-70, 1991.

[9] B. Moghaddam and A. Pentland, “Probabilistic Visual Learning for Object Recognition”, IEEE Transaction on Pattern Analysis an Machine Intelligence, vol. 19, no. 7, pp. 696-710, July 1997.

[10] H. Murase and S. K. Nayar, “Visual Learning and Recognition of 3D Objects from Appearance”, International Journal of Computer Vision, vol.14, pp. 5-24, 1995.

[11] K. Ohba and K. Ikeuchi, “Detectability, Uniqueness, and Reliability of Eigen Windows for Stable Verification of Partially Occluded Object”, IEEE Transaction on Pattern Analysis an Machine Intelligence, vol. 19, no. 9, pp. 1043-1048, September 1997.

[12] A. Papoulis, “Probability, Random Variables, and Stochastic Processes”, 3rd Edition McGraw-Hill, 1991.

[13] W. H. Press, W. T. Vetterling, S. A. Teukolsky and B. P. Flannery, “Numerical Recipes in C. The art of scientific computing”, 2nd edition Cambridge University Press, 1992.

[14] S. Ravela and R. Manmatha , “Retrieving images by appearance”, Proc. of the Sixth International Conf. on Computer Vision (ICCV 98), pp.608-613, Bombay, India, January 1998.

[15] P. J. Rousseuw, “Least Median of Squares Regression”, Journal of the American Statistical Association, vol. 79, no. 388, pp. 871-880, December 1984.

[16] B. Schiele, J.L. Crowley, “Probabilistic object recognition using multidimensional receptive field histograms”, Proc. of Fourth European Conf. on Computer Vision (ECCV 96), vol. 1, pp.610-619, April 1996.

[17] C. Schmid, R. Mohr, “Local Grayvalue Invariants for Image Retrieval”, IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 19, no. 5, pp.530-535, May 1997.

[18] C. V. Stewart, “MINPRAN: A New Robust Estimator for Computer Vision”, IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 17, no. 10, pp. 925-938, October 1995.

[19] C. V. Stewart, “Bias in Robust Estimation Caused by Discontinuities and Multiple Structures”, IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 19, no. 8, pp.818-833, August 1997.

[20] L. Zhao and Y.-H. Yang, ”Mosaic Image Method: a Local and Global Method” , Technical Report, University of Saskachewan, Canada, June 1997.