SIMULTANEOUS RECONSTRUCTION AND RECOGNITION OF NOISY CHARACTER-LIKE SYMBOLS

László Czúni[1], Ágnes Lipovits[2] and Dávid Papp[1]
[1]*Dept. of Electrical Engineering and Information Systems, University of Pannonia, Egyetem street 10., Veszprém, Hungary*
[2]*Dept. of Mathematics, University of Pannonia, Egyetem street 10., Veszprém, Hungary*

Keywords: Image Reconstruction, Markov Random Field, Optical Character Recognition, Radon Transform

Abstract: In our article we deal with the simultaneous problem of reconstruction and recognition of binary symbols loaded with heavy additive noise. We introduce a Markov Random Field (MRF) model where a shape energy term is responsible to find a solution similar to a tested hypothesis. This way we could increase the precision of the reconstruction process the only question is how to find out the right hypotheses which helps the reconstruction the best way. Fortunately the new energy term gives us the answer: the tested hypotheses with the minimal shape energy component designates the right shape.

1 INTRODUCTION

The reconstruction of binary images from noisy observations is a common problem in image processing. In several applications, besides observations, we have some a priori information about the possible shapes of the typical objects (such as letters or other well defined symbols) to help the reconstruction. Often we would also like to recognize the observed symbols, not possible without prior noise filtering.
In our paper we show an MRF based solution for this twofold problem: the proposed algorithm makes reconstruction suitable for visual purposes or for further processing, and it also recognizes the symbols. For a priori information about the shape of objects, their Radon transform is stored as simple shape descriptor vectors. The Radon transform is fast to compute, requires only 1D information to store, and tolerates some distortion of the original shapes.

2 THE MRF MODEL

Markov Random Field (MRF) is a probability model based on local characteristics (Geman and Graffigne, 1986). In an MRF, the sites are related to one another via a neighborhood system. A generalized definition of Markov Random Field can be given using graphs. Let $G=(S,E)$ be an undirected graph where $S=s_1,s_2,\dots,s_N$ is a finite set of vertices (sites) of the graph, and E is the set of edges of the graph. By definition, two sites of the graph, s_i and s_j are neighbors if there is an edge connecting them. Given a site s, the set of points which are neighbors of s (the neighborhood of s) is denoted by V_s. By definition, $V=\{V_s|s\in S\}$ is a neighborhood system for G if

$$s\notin V_s \text{ and } s\in V_r \Leftrightarrow r\in V_s \quad (1)$$

We assign to each site of the graph a label λ from a finite set of labels Λ. Such an assignment is called a configuration, denoted by ω. The set of all possible configurations is denoted by Ω. The configuration restricted to a subset $T\subset S$ is denoted by ω_T. The value given to a site s by the configuration ω is represented by ω_s. We assign probability measures to the set Ω of all possible configurations ω. The local characteristics of a probability measure P defined on the set Ω of all possible configurations are the conditional probabilities of the form $P(\omega_s|\omega_{S-s})$, that is, the probability that the site s is assigned the label ω_s, given the values at all other sites of the graph. By definition, a probability measure χ is a Markov Random Field with respect to a neighborhood system V if

$$\forall\omega\in\Omega: P(\chi=\omega)>0 \quad (2)$$

$$\forall s\in S,\forall\omega\in\Omega: P(\omega_s|\omega_{S-s})=P(\omega_s|\omega_V) \quad (3)$$

so that the local characteristics of the probability measure depend only on the knowledge of the labels at the neighboring sites.

By definition, $C\subset S$ is a clique, if every pair of points in C are neighbors. We can define a potential V as a way to assign a number $V_a(\omega)$ to every subconfiguration ω_A of a configuration ω. Given the potential

Czúni L., Lipovits Á. and Papp D..
SIMULTANEOUS RECONSTRUCTION AND RECOGNITION OF NOISY CHARACTER-LIKE SYMBOLS.
DOI: 10.5220/0003861101970200
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2012), pages 197-200
ISBN: 978-989-8565-03-7

V, it defines an energy $U(\omega)$ on the set Ω of all configurations ω by $U(\omega) = -\sum_A V_A(\omega)$ where for a fixed ω the sum is taken over all subsets A of S.

The Gibbs measure induced by U is defined by:

$$\pi(\omega) = \frac{1}{Z} exp(-U(\omega)), Z = \sum_{\omega} exp(-U(\omega)). \quad (4)$$

In the case when $V_A(\omega) = 0$ whenever A is not a clique, the potential V is called a nearest neighbor Gibbs potential. The Hammersley-Clifford theorem establishes the equivalence between Gibbs measures and MRFs: χ is an MRF with respect to the neighborhood system V, if and only if $\pi(\omega) = P(\chi = \omega)$ is a Gibbs distribution with a nearest neighbor Gibbs potential V, that is

$$\pi(\omega) = \frac{1}{Z} exp(-\sum_{c \in C} V_c(\omega)), \quad (5)$$

where C is the set of cliques. This equivalence allows us to specify potentials instead of local characteristics when defining an MRF.
Markov Random Fields can be used to address many low-level image tasks, like restoration, edge detection, segmentation, motion detection. Let F denote the observations on a grayscale image, and f_s denote the observation belonging to the pixel s. We have to find the configuration ω which maximizes the probability $P(\omega|F)$. By the Bayes-theorem

$$P(\omega|F) = \frac{1}{P(F)} P(F|\omega)P(\omega). \quad (6)$$

Furthermore,

$$P(F|\omega) = \prod_{s \in S} P(f_s|\omega_s), \quad (7)$$

and based on the Hammersley-Clifford theorem,

$$P(\omega) = \prod_{c \in C} exp(-U_c(\omega_c)), \quad (8)$$

where C is the set of all possible cliques, and U is the clique potential. So we have to find the configuration which maximizes the function

$$\prod_{s \in S} P(f_s|\omega_s) \prod_{c \in C} exp(-U_c(\omega_c)) \quad (9)$$

$P(F)$ can be omitted, because it does not depend on ω. Assuming that $P(f_s|\omega_s)$ is Gaussian, and taking the logarithm of the above we get the following energy functions:

$$U_1(\omega, F) = \sum_{s \in S} (\ln \sqrt{2\pi}\sigma_{\omega_s} + \frac{(f_s - \mu_{\omega_s})^2}{2\sigma^2_{\omega_s}}) \quad (10)$$

which stands for the observation, and

$$U_2(\omega) = \sum_{c \in C} V_2(\omega_c) \quad (11)$$

$$V_2(\omega_c) = V_{s,r}(\omega_s, \omega_r) = \begin{cases} -\beta \; if \; \omega_s = \omega_r \\ +\beta \; if \; \omega_s \neq \omega_r \end{cases} \quad (12)$$

β is a model parameter representing the homogeneity of the regions, μ_{ω_s} and σ_{ω_s} is the mean and the deviation belonging to label ω_s. The goal is to find a configuration which minimizes the two energy functions,

$$\sum_{s \in S} V_1(\omega_s, f_s) + \sum_{c \in C} V_2(\omega_c) \quad (13)$$

For optimization different relaxation techniques can be used, in our paper we apply Simulated Annealing.

2.1 Radon Transformation

Radon transformation is widely used in the field of tomography (Radon, 1917) (Nagy and Kuba, 2005) (Naser et al., 2009). The projection of an object at a given angle θ is made up of a set of line integrals. These line integrals are the Radon transform of the object. The inverse Radon transform can be used to reconstruct an approximation of the original object. If we had an infinite number of projections of an object taken at an infinite number of angles, we could perfectly reconstruct the original object.

Let $f(x,y)$ be a two dimensional continuous function. The Radon transform is a function defined on the two dimensional space of straight lines L by the line integral along each line:

$$Rf(L) = \int_L f(x)dx \quad (14)$$

A straight line L can be represented in the form:

$$(\rho\cos\theta - s\sin\theta, \rho\sin\theta + s\cos\theta), \quad (15)$$

where ρ is the distance of L from the origin, and θ is the angle of the normal vector to L with the x axis.

The Radon transform can be expressed, using these quantities as coordinates, as:

$$Rf(\rho, \theta) = \int_{-\infty}^{\infty} f(\rho\cos\theta - s\sin\theta, \rho\sin\theta + s\cos\theta)ds \quad (16)$$

We use only two angles (the vertical and horizontal directions), which makes it easy and fast to compute. Let $f(i,j)$ be a binarized image, where

$$f(i,j) = \begin{cases} 1 & \textit{if the pixel is black} \\ 0 & \textit{if the pixel is not black} \end{cases} \quad (17)$$

$0 \leq i < n, 0 \leq j < m$, where n is the height of the image, and m is the width of the image.

We denote the Radon transform of the jth row as Rr_j, and the Radon transform of the ith column as Rc_i.

We define them as the sum of the values of the pixel in the jth row or the ith column:

$$Rr_j = \sum_{i=0}^{n-1} f(i,j),\ Rc_i = \sum_{j=0}^{m-1} f(i,j). \tag{18}$$

We can define the probability of a random pixel chosen from a row or a column being black, in respect of the chosen row or column:

$$P(r_j(i)=1) = \frac{Rr_j}{n} \text{ and } P(c_i(j)=1) = \frac{Rc_i}{m}. \tag{19}$$

Considering the whole image, the probability is the following:

$$P(f(i,j)=1) = P(c_i(j)=1) \cdot P(r_j(i)=1). \tag{20}$$

2.2 The Proposed Extension of MRF

For the MRF reconstruction with Radon transforms (called mMRF), a hypothesis about the shape is needed since we will have a new energy term to measure the similarity to this hypothesis. For simplicity we run the mMRF with all of the possible symbols as hypotheses, and then we find out the right one as the reconstruction is finished. The similarity between the reconstruction and the hypothesis is described by the energy term:

$$U_3 = \gamma \cdot \sum_{i=0}^{n-1}\sum_{j=0}^{m-1} |P(r_j(f,i)=1) - P(r_j(g,i)=1)| + |P(c_i(f,j)=1) - P(c_i(g,j)=1)|, \tag{21}$$

where γ is a constant specifying a weight, f is the reconstructed image, and g is the hypothesis image.

2.3 Localization

The position of the characters under reconstruction is not known but would be required according to (21). The vertical coordinate can be easily determined since the characters are usually part of a bigger textual data, where the position of the lines can easily be found. To calculate the horizontal coordinate matching of Radon transforms can be used. Unfortunately, the Radon functions of the raw input images are too noisy (see Figure 2), so we apply a Gaussian convolution filter on the input image and then binarize the result. Matching is done by:

$$\min_x \sum_{i=0}^{n-1} |Rc_{i+x}(f) - Rc_i(g)|, \tag{22}$$

where g is the smoothed and binarized noisy image, and f is the original template image.

Table 1. shows, for different font types, the percentage of correct localization of the horizontal position and the average distances from the real position in pixels in our experiments. Even if the estimated position is not precise, it is at maximum 2 pixels away from the real position and the average difference is at subpixel level, it will not affect the classification seriously, as shown later.

Table 1: Results of the character localization using Radon transforms after preprocessing.

Font type	Exact match	Average distance
arial	83,33%	0.1667
kristen	83,33%	0.1667
pescadero	66,67%	0.3611

3 RECONSTRUCTION AND CLASSIFICATION

In our experiments three different fonts were used, each of 36 characters (the alphabet from a to z and the digits from 0 to 9). Figure 1. shows samples of some of the tested characters with and without noise (Gaussian additive noise with 0 mean and 400 variance was applied).

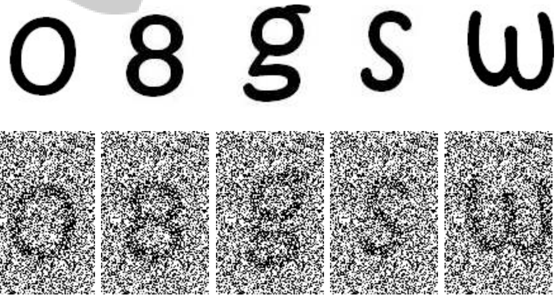

Figure 1: A sample of the kristen characters without and with noise.

It is very common to recognize characters by using only their Radon transforms (Miciak, 2010). The row and the column Radon transform could be seen as vectors, and the Euclidean distance of these vectors could be used as a measure to find the closest match to the noisy image. Unfortunately, due to the very heavy noise, the Radon transforms can not provide a method for proper recognition.

Figure 2. shows the Radon transforms of two characters with and without noise. The noise-free versions can be easily distinguished, however, the ones with the heavy noise do not have any unique characteristics, they are very similar to each other. They can not be recognized by simply using the distance of the

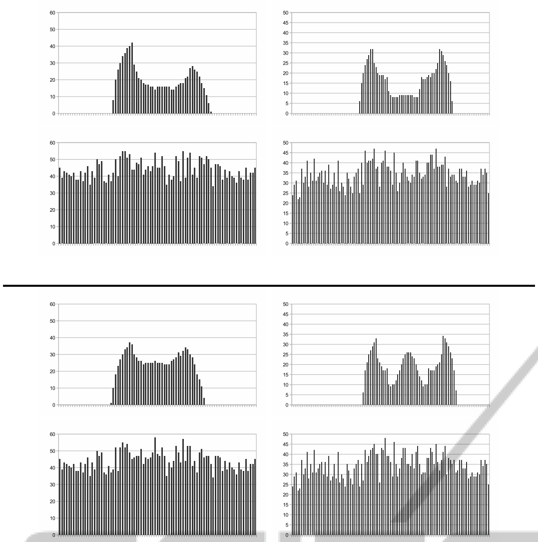

Figure 2: The vertical (left) and the horizontal (right) Radon transforms of the noise-free and noisy image of the character 'c' (above line) and character 's' (below line).

Radon transforms as experienced in our tests, where only 1 (of the 36) character was correctly classified. The new idea of our proposal is to use the Radon functions' similarity in the reconstruction and let this energy term (21) decide if the hypothesis is correct or not. We chose the hypothesis with the minimal U_3 to be the right one. Thus we assume that the MRF process will result in the best reconstruction among others and will also classify the symbol by U_3

Figure 3. shows samples of the images reconstructed with MRF, and with mMRF. Table 2. shows the percentage of improvement of reconstruction and the classification rate for each font using the proposed method.

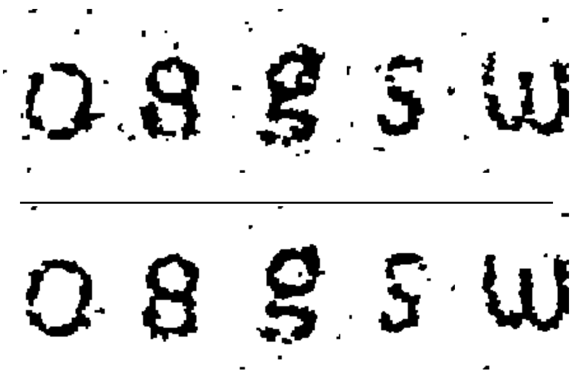

Figure 3: A sample of kristen characters reconstructed with MRF and mMRF.

Table 2: Improvement of reconstruction rate over the normal MRF and percentage of successful classifications.

	arial	kristen	pescadero
Classification	100,00%	100,00%	91,67%
Reconstruction	9.97%	8,25%	13,72%

4 SUMMARY

In our paper we investigated the reconstruction of binary symbols with very low SNR. We found that with the proposed extension of the plain MRF model the pixel-based reconstruction increased with appr. 11% in general in case of three types of test fonts. We found that the right template, necessary for reconstruction, can be designated by the new energy term generated with the Radon transform. Classification rate was 97% for the tested three font types.

ACKNOWLEDGEMENTS

The work and publication of results have been supported by the Hungarian Research Fund, grant OTKA CNK 80368.

REFERENCES

Geman, S. and Graffigne, C. (1986). Markov random field image models and their applications to computer vision. In *Proceedings of the International Congress of Mathematicans*. American Mathematical Society.

Miciak, M. (2010). Radon transformation and principal component analysis method applied in postal address recognition task. *International Journal of Computer Science and Applications*, 7(3):33–44.

Nagy, A. and Kuba, A. (2005). Reconstruction of binary matrices from fan-beam projections. *Acta Cybernetica*.

Naser, M. A., Mahmud, A., Arefin, T. M., Sarowar, G., and Ali, M. M. N. (2009). Comparative analysis of radon and fan-beam based feature extraction techniques for bangla character recognition. *IJCSNS International Journal of Computer Science and Network Security*, 9(9):287–289.

Radon, J. (1917). ber die bestimmung von funktionen durch ihre integralwerte lngs gewisser mannigfaltigkeiten. *Berichte ber die Verhandlungen der Schsische Akademie der Wissenschaften*, pages 262–277.