ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 89-94

Competitive Neural Networks for robust computation of optical flow

Elsa Fernandez, Imanol Echave, Manuel Graña
Dept CCIA, UPV/EHU,
ccpgrrom@si.ehu.es

Abstract. The Self Organizing Map and the Simple Competitive Learning are used to compute adaptively the vector quantizers of color image sequences. The codebook computed for each image in the sequence is then used as a smoothing filter, the VQ Bayesian Filter (VQ-BF), for the preprocessing of the images in the sequence. The optical flow is then robustly and efficiently computed over the filtered images applying a correlation method on the isolated pixels.

1. Introduction

The computation of the optical flow is a central issue for many of the tasks that are involved in the application of computer vision to robotics[5], [7], [10]. In this paper we propose a correlation based approach which computes the optical flow estimates upon a preprocessed image sequence. We propose the filtering of the image sequence through the application of adaptive color quantization. The color quantization involves the consideration of spatial neighborhoods to regularise and smooth the optical field. We call our approach Vector Quantization Bayesian Filter (VQ-BF). Each pixel is coded according to the vector quantization of its neighborhood. The codebook is computed adaptively by Competitive Neural Networks, i.e.: the Self Organizing Map [11]. For lack of space the results of the experiments are not included in the paper, they can be found in the following web address: http//sizx01.si.ehu.es/imanol/esann2000/. In these experiments, the SOM is used to compute the codebook over the first image in the sequence. Afterwards, a Simple Competitive Learning algorithm (the SOM with null neighborhood) is applied to each image in the sequence to adapt the codebook to changing illumination conditions and color distributions. The VQ-BF approach produces a smoothing of the images which preserves the boundaries. Such kind of smoothing reduces the spurious flow due to illumination, microtexteures and noise, while the preservation of the boundaries produces enhanced estimations at the critical boundaries. These qualities of the VQ-BF smoothing allow for the computation based on single pixels of the correlation between images.

The paper is structured as follows: Section 2 presents our application of the codebooks as a filtering mechanism, section 3 comments the one-pass SOM

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 89-94

applied to extract the codebook, section 4 presents the correlation method for optical flow computation, section 5 presents the experimental results that can be found in the referred page. Finally, section 6 gives some conclusions and further work.

2. Vector Quantization as a Bayesian Filter (VQ-BF)

We start recalling some standard notation [12]. An image will be described by an array $x = \left(x^P, x^L\right)$ Let S^P denote a finite square lattice, each representing a pixel position. The vector $x^P = \left(x_s^P\right)_{s \in S^P}$ represents a pattern or configuration of grey values. In texture classification a pattern of labels is represented by $x^L = \left(x_s^L\right)_{s \in S^L}$ where S^L is a set of pixel blocks and $x_s^L = l \in L$ is the label of block s. The blocks may be overlapping. The observable data y is a function Y of the true image x. The conditional law of Y will be denoted by $\mathsf{P}\left(y \left| x\right.\right)$. The prior Π and the conditional probabilities determine the joint distribution of data and images by $\mathsf{P}\left(x, y\right) = \Pi\left(x\right) \mathsf{P}\left(y \left| x\right.\right)$. The posterior probability is given by Bayes' theorem $\mathsf{P}\left(x \left| y\right.\right) = \Pi\left(x\right) \mathsf{P}\left(y \left| x\right.\right) \left[\sum_z \Pi\left(z\right) \mathsf{P}\left(y \left| z\right.\right)\right]^{-1}$. For continuous data, the prior distribution Π will in general have the Gibbsian form $\Pi\left(x\right) = Z^{-1} \exp\left(-H\left(x\right)\right)$, with $Z = \sum_{z \in \mathbf{X}} \exp\left(-H\left(z\right)\right)$. Here H is the energy function of Π. In most cases, the posterior is again of Gibbsian form. A mode of the posterior distribution $\widehat{x} = \max_x \left\{\mathsf{P}\left(x \left| y\right.\right)\right\}$ is called a maximum a posteriori (MAP) estimate of x given y. For Gibbsian posterior distributions the MAP corresponds to the minimum energy. Let the observation be of the form $Y = \varphi\left(X, \eta\right)$.The law of the noise η is denoted by Γ. If the noise η and the image X are independent then the conditional probabilities are of the form $\mathsf{P}\left(Y = y \left| X = x\right.\right) = \Gamma\left(\varphi\left(X, \eta\right) = y\right)$. In particular, if the noise is Gaussian, the conditional probabilities and the posterior probabilities can be written in a Gibbsian form [3], [12].

2.1. The model for the image blocks

Let us consider that we have a set of image block representatives, i.e. a codebook $\Omega^* = \left\{\omega_i^*; i = 1, .., c\right\}$ where each representative is an image of size $d \times d$, $\omega_i = \left(x_s^P\right)_{s \in S^{d \times d}}$ with the lattice $S^{d \times d}$ defined as a neighborhood $S^{N(d \times d)} = \left\{s : -d/2 \leq |s| \leq d/2\right\}$. Given a sample of observed image blocks $\mathbf{y}^b = \left\{y_i^b; i = 1, .., n;\right\}$.with $y_i^b = \left(y_s^P\right)_{s \in S^{d \times d}}$, the codebook is the result of a Vector Quantization design algorithm that tries to minimize some objective function. Let us consider that the codebook is designed to minimize the mean square error over the sample $\Omega^* = \min_\Omega \left\{\sum_{i=1}^n \left\|y_i^b - \omega_{j(i)}\right\|^2\right\}$where $j\left(i\right) = y\left(y_i^b\right) = \arg\min\left\{\left\|y_i^b - \omega_j\right\|^2; j = 1, .., c\right\}$. This minimization corresponds to the maximum likelihood estimation of the parameters (the class means) of a mixture of Gaussians of identity covariance matrices, which is

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 89-94

the assumed model of the distribution of the observed image blocks. We assume equal prior probabilities of the image blocks . The VQ decision given by $\widehat{x} = \omega^*_{j(y^b)}$ corresponds to a MAP decision assuming posterior probabilities of the image blocks x^b could be put into the Gibbsian form with energy functions: $H\left(Y^b = y^b \left| x^b = \omega^*_j \right.\right) = H\left(x^b = \omega^*_i \left| Y^b = y^b \right.\right) = \frac{1}{2}\left\|y^b - \omega^*_i\right\|^2$. The consideration of objective functions other than the distortion would lead to other probabilistic models.

2.2. The model for the VQ-BF

In the VQ-Bayesian Filter (VQ-BF) the image is not decomposed into blocks. For each pixel $s \in S^P$ we select a window around it $S_s^{N(d\times d)} \subset S^P$. Each image window $y^b_s = \left(y^P_{s'}\right)_{s' \in S_s^{N(d\times d)}}$ is processed independently. We call the approach Bayesian because the pixel process is conditioned on its neighborhood. In classification mode each pixel is classified according to its neighborhood:

$$\widehat{x}^L_s = j\left(y^b_s\right), s \in S^P \tag{1}$$

If we consider independent the pixel neighborhoods, the posterior probabilities are Gibbisan with a posterior energy of the form: $H\left(x\left|y\right.\right) = C + \frac{1}{2}\sum_s \left\|y^b_s - \omega^*_{j(y^b_s)}\right\|^2$. The prior corresponds to the joint probability of the pixel windows $\Pi\left(X = x\right) = P\left(x^b_s = \omega^*_s; s \in S^P\right)$and can not be easily put in product form. An approximation to its Gibbsian form may be the assumption that neighboring pixels will have the same class value if the variation of their surrounding windows is small. This is a smoothing prior that depends on the codebook. Taking into account the relation [12] $\ln \mathsf{P}\left(y\left|x\right.\right) = C - H\left(\widehat{x}\left|\widehat{y}\right.\right) - H(x)$, and assuming prior energy $H(x) = \sum_s \sum_{t\in N(s)} \left(x_s - x_t\right)^2 \left\|x^b_s - x^b_t\right\|^2$ we arrive to the following approximate expression for the log-likelihood

$$\ln \mathsf{P}\left(y\left|x\right.\right) \approx -\sum_s \left\|y^b_s - \omega^*_{j(y^b_s)}\right\|^2 - \sum_s \sum_{t\in N(s)} \left(x_s - x_t\right)^2 \left\|x^b_s - x^b_t\right\|^2 \tag{2}$$

If we consider the Gaussian additive noise, then the expression for the deformations that are deduced from the likelihood are:

$$Y = \varphi\left(X,\eta\right) = \widehat{\eta} - \left(\sum_s \left\|y^b_s - \omega^*_{j(y^b_s)}\right\|^2 - \widehat{\eta}\right) - \sum_s \sum_{t\in N(s)} \left(x_s - x_t\right)^2 \left\|x^b_s - x^b_t\right\|^2$$

This expression can be interpreted as stating the ability of VQ-BF to recover from smooth deformations involving the pixel's neighborhoods.

3. The Self Organizing Map

A critical step in the application of our approach is the computation of the codebook, because it must be representative of the true statistics in the image.

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 89-94

As we have discussed, these statistics are the building blocks for the priors that underly VQ-BF. Also, the compuutation must be done in a robust an fast way. We apply a one-pass Self Organizing Map [11] for this task, assuming it as a robust initialization for the steepest gradient descent of the Euclidean distortion [6].

Let it be $\mathbf{y} = \{y_i; i = 1, .., n; \}$ be a sample of vectors (image blocks), that may be an infinite source $n = \infty$, and let $\Omega = \{\omega_i; i = 1, .., c\}$ denote the codebook we are computing. The Self Organizing Map is an online algorithm that computes the following rule after the presentation of a new input y_t from the sample at time t, $\triangle\omega_{i,t} = \alpha_{i,t} \upsilon_{i,t} (y_t, \Omega_t) (y_t - \omega_{i,t})$where $\upsilon_{i,t} (y_t, \Omega_t)$ is a neighboring function that depends on the topology defined on the indices of the codevectors, $\alpha_{i,t}$ is the learning rate that decreases to zero in the usual fashion for the stochastic gradient approach. This neighboring function usually is defined as decreasing with the distance and shrinking as the adaptation time proceeds. In our experiments we have set the learning rate to $\alpha_{i,t} = 0.1 (1 - \tau_i/n)$ with $\tau_i = \sum_{k=1}^{t} \upsilon_{i,k} (y_k, \Omega_k)$, so that the learning rate will be sensitive to the size of the cluster associated with each codevector. The neighboring function is of the form $\upsilon_{i,t} (y_t, \Omega_t) = \{1 \text{ if } |i - j (y_t)| \leq v_t; 0 \text{ otherwise}\}$. The radius v_t of neighboring function follows the expression $v_t = \left\lceil (v_0 + 1)^{\left(1 - \frac{r}{n} t\right)} \right\rceil - 1$ for $t < \frac{n}{r}$. For $t > \frac{n}{r}$ we assume that the SOM behaves like the Simple Competitive Learning algorithm and the neighboring function becomes the hard clustering criterium. The sample is presented only once.

4. Correlation approach to the optical flow

The computation of the optical flow [5] based on the correlation of image patches between frames in the image sequence has been presented in [1], [8], in [7], [10] it is presented under the framework of image matching. The basic algorithm consists in the matching of each pixel and its neighborhood of size $v \times v$ in the next frame inside a radius of maximum movement η. For each pixel (x, y) and shift (u, w) inside the allowed radius of movement $|u|, |w| \leq \eta$. The matching that gives the likelihood that the pixel has moved along the vector (u, w) while passing from image $E_t (x, y)$ to image $E_{t+1} (x, y)$ is of the form $M_{t,t+1} (x, y; u, w) = \sum_{i,j=-v}^{v} \phi (E_t (x + i, y + j), E_{t+1} (x + i + u, y + j + w))$where $\phi (a, b) = |a - b|$ is the matching function. The motion vector at pixel (x, y) : $F_{t,t+1} (x, y)$ is determined as $F_{t,t+1} (x, y) = \arg\min_{-\eta \leq u,w \leq \eta} \{M_{t,t+1} (x, y; u, w)\}$. It can be deduced from this expression that the algorithm complexity is of order $O (N v \eta)$.

5. Experiments and results

We have applied our approach to several image sequences in the following way: (1) The Self Organizing Map is applied to the first image in the sequence to

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 89-94

obtain the initial codebook. The Simple Competitive Learning is used afterwards for small adaptations of the codebook at each image. (2) The images in the sequence are then filtered with the VQ-BF approach using this codebook, and the optical flow is computed at the pixel level. That is, the correlation is computed with neighborhood width $v = 0.$ and maximum radius movement $\eta = 3$. The codebook size is 4 and the dimension of the codevectors is 5x5. A prototype of the algorithms has been made in IDL. Its response time is in the order of seconds for each frame. We expect the optimized C implementation to become near real time.

The reader can find in http://sizx01.si.ehu.es/imanol/esann2000/ the experimental results given by the filtered sequences and the optical flow. In general the smoothing produced by the VQ-BF produces null optical flow estimations in constant, or near constant surfaces with some microtexture like walls or the doors. Although no boundary detection is performed the significative boundaries are the key elements in the resulting flow. The sequences shown in the web site are a zooming, a translation sequence, and a couple of people moving in front of the camera. The zooming sequence simulates a translational movement in the direction of the optical axis. This the kind of image sequence that could correspond to a vision based collision detection process. The smoothing reduces much of the effects of the uncontrolled illumination.. The effect of the clutter is minimal. The translation sequence corresponds to a rotation of the camera around the x axis of the image plane. The effect is a panning that could correspond either to a search in the scene or to an obstacle avoidance manoeuver.. The wall shows some phantom flows due to illumination gradients. These gradients are preserved in the filtering because of the big samples involved by the color variations in the wall force the assignment of different clusters by the SOM. This is the main disadvantage of our approach: the oversegmentation due to the imbalance of the data extracted from the images. The effect is the presence of phantom boundaries in very large regions with small gradients.

6. Conclusions and further work

Based on the adaptive computation of the codebooks for image VQ done with Competitive Neural Networks, we propose a preprocessing stage for the computation of the optical flow. This preprocess consists of a smoothing of the images based on the vector quantization of the pixel neighborhoods. The smoothed images are used to compute the optical flow with a correlation method applied to the isolated processed piexels. The smoothing preserves the boundaries improving the estimation of the flow at the boundaries while eliminating or reducing the spurious detections due to noise effects in the homogeneous regions. The experiments show the benefits of our approach. The optical flow computation is intended for the segmentation and isolation of human figures in arbitrary environments. The intended application is the detection by mobile robots of human figures based on active vision and the processing of color. This

ESANN'2000 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 26-28 April 2000, D-Facto public., ISBN 2-930307-00-5, pp. 89-94

detection is the first step for the robots to navigate in human populated environments, start interactions such as identification and interpretation of body language to avoid collisions..

Acknowledgments Work supported by projects PI1998/21 and TAP98-0294-c02-02. Imanol Echave has a predoctoral grant from the Dept. de Industria of the Gobierno Vasco.

References

[1] T. Camus (1997) "Real time quantized optical flow" Real Time Imaging 3(2):71-80

[2] Cosman P.C, Oehler K.L., Riskin E.A., Gray R.M. (1993) "Using Vector Quantization for Image Processing" IEEE Proceedings 81(4):1326-1341

[3] Geman S., Geman D. (1984) Stochastic relaxation, Gibbs Distributions and the Bayesian Restoration of Images IEEE trans. PAMI 6:721-741

[4] Gersho A., Gray R.M. (1992) Vector Quantization and Signal Compression New York: Kluwer

[5] BKP Horn "Robot Vision" MIT Press (1986) (1990)

[6] Gonzalez A. I., Graña M., D´Anjou A., Albizuri F.X., Cottrell M. (1997) A sensitivity analysis of the Self Organizing Map as an Adaptive One-pass Non-stationary Clustering algorithm: the case of Color Quantization of image sequences Neural Processing Letter 6:77-89

[7] R.M. Haralick, L.G. Shapiro (1993) "Computer and robot vision" Reading, MA: Addison-Wesley

[8] B. Jahne (1993) "Spatio-Temporal image processing. Theory and Scientific applications". Berlin:Springer Verlag, LNCS 751

[9] Jain A.K. (1989) "Fundamentals of digital image processing" Englewood-Clliffs: Prentice-Hall

[10] R. Jain, R. Kasturi, B.G. Schunck (1995) " Machine Vision" New York: MacGraw Hill

[11] Kohonen T. (1995) "Self Organizing Maps" Berlin: Springer-Verlag

[12] Winkler G. (1995) "Image analysis, random fields and dynamic Monte Carlo methods" Berlin: Springer Verlag