USING PERCEPTRON-LIKE ALGORITHMS FOR THE ANALYSIS AND THE PARAMETERIZATION OF OBJECT MOTION

Marco Mattavelli§ and Edoardo Amaldi†[1]
§Signal Processing Laboratory, Swiss Federal Institute of Technology,
CH-1015 Lausanne, Switzerland
marco@lts.de.epfl.ch
†School of Operations Research and Center for Applied Mathematics,
Cornell University, Ithaca, NY 14853, USA
amaldi@cs.cornell.edu

ABSTRACT

A new approach based on the extraction of maximum consistent subsystems of linear systems is proposed for addressing the general problem of determining the linear motion parameters of unknown moving objects from a sequence of images. This type of task can be tackled using simple but effective variants of the well-known perceptron algorithm that aim at maximizing the number of patterns that are correctly classified. Unlike in the usual perceptron applications, the weight vectors determined during the training phase are not used to classify new patterns but to extract the structure and to provide the parameters of the considered piecewise linear model. The potentialities of the new approach are demonstrated for the segmentation of the optical flow. Experimental results obtained for fields from synthetic and natural images indicate various advantages of our approach with respect to some classical alternatives.

1. INTRODUCTION

The extraction and interpretation of the motion of the sensor and of independently moving objects from visual information is a fundamental problem in image processing and computer vision. Two classes of approaches have been proposed in the literature: one is based on the computation and interpretation of the optical flow [1, 2, 3, 4] while the other extracts motion parameters directly from the image sequences and leads to a layered representation of object motions [5, 6]. The major problem of the first class of approaches is how to combine local features in order to extract global object motions. Since flow fields are usually noisy and partially incorrect, especially near occlusions and motion boundaries, global approaches are considered necessary in order

[1] E. A. is partially supported by a Postdoctoral fellowship of the Swiss National Science Foundation.

to yield robust parameter estimates. But the delicate choice of the objective function and the presence of several moving objects make these techniques complex and unsatisfactory. For the second class of approaches, the difficulty lies in partitioning the data into a consistent layered representation. Unless the segmentation is already correct, a set of pixels being fitted may contain multiple motions instead of a single dominant coherent component.
Methods that combine and iterate the two types of techniques have also been proposed in order to achieve more accurate representations and interpretations of the optical flow [7]. A common difficulty of both classes of methods is the *chicken and egg problem* of the parameter and structure estimation. If the structure is known, it is relatively easy to correctly estimate the parameters. But if the parameters are not known it is difficult to correctly extract the structure. This kind of dilemma, which pertains to several *ill-posed* problems, assumes a fundamental importance in the analysis and interpretation of motion data.
In this paper we address the problem of determining an unknown linear motion structure (several independent moving objects) with unknown parameters from a sequence of images. A new approach based on the extraction of maximum size consistent subsystems of linear systems is proposed for the segmentation of the optical flow. The reported results indicate that a simple perceptron-like algorithm can efficiently extract the motion parameters and the motion structure without requiring a dominant motion or any *a priori* hypothesis on the number and on the spatial location of the moving objects.

The paper is organized as follows. Next section describes how the optical flow segmentation problem can be formulated in terms of finding maximum consistent subsystems of linear systems. In section 3 we briefly recall the thermal variant of the perceptron algorithm [8, 9] and show how it can be used to determine maximum consistent subsystems of linear equations. Section 4 points out the major shortcomings of alternative approaches such as robust estimation [10, 11] and the Hough transform [12, 13]. A few typical experimental results on synthetic optical flow fields are reported in section 5. Finally, some concluding remarks are mentioned in section 6.

2. PROBLEM FORMULATION

Consider an image sequence that evolves in time with the apparent motion given by equation:

$$I(x,y,t) = I(x - d_x(x,y), y - d_y(x,y), t-1), \tag{1}$$

where x and y are the image coordinates, $I(\cdot)$ is the image intensity and $d_x(x,y)$ and $d_y(x,y)$ denote the image displacements at each point. When the distances between the object surfaces and the camera are large, it is usually possible to approximate the motion as an affine transformation:

$$d_x(x,y) \quad = \quad w_1 + w_2 x + w_3 y \tag{2}$$

$$d_y(x,y) \quad = \quad w_4 + w_5 x + w_6 y$$

for every image point (x, y).
Let us assume that the optical flow has been estimated using a standard technique [1], i.e. we have a couple of components $(d_x(x_k, y_k), d_y(x_k, y_k))$ for n image points (x_k, y_k) with $k \in \{1, 2, ..., n\}$. The problem of the interpretation of the optical flow arising from object motion is to segment the field into an unknown number r of regions that correspond to the same affine motion and to estimate the corresponding set of parameters $w_1^l....w_6^l$, where $l \in \{1, 2, ..., r\}$. Consider the linear system consisting of the n pairs of linear equations (3) in the six variables $w_1....w_6$. For multiple motions and reliable optical flow estimates, the overall system is inconsistent and the regions with the same affine motion correspond to consistent subsystems. Formally, the optical flow segmentation problem can be formulated as that of partitioning the system of $2n$ equations (3) into r disjoint consistent subsystems and of finding a set of corresponding solutions $w_1^l, ...w_6^l$ with $l \in 1, 2, ..., r$. Since all pairs of points with the same affine motion should be associated to the same region, one looks for maximum consistent subsystems, i.e. subsystems containing the largest number of pairs of equations (3). In order to partition the image into a small number of regions, a simple greedy strategy is adopted in which as large as possible consistent subsystems are removed iteratively.

3. PERCEPTRON-LIKE ALGORITHMS AND OPTICAL FLOW SEGMENTATION

A *perceptron* is a simple linear threshold unit whose output

$$y = \begin{cases} +1 & \text{if } \sum_{j=1}^{n} w_j x_j \geq 0 \\ -1 & \text{otherwise,} \end{cases}$$

where $x_j \in \Re$ denotes the jth input and $w_j \in \Re$ the jth weight. Given a task, i.e. a set of p input vectors $\mathbf{a}^k$, $1 \leq k \leq p$, and the corresponding desired output $b^k \in \{-1, 1\}$, the objective of *training* is to find a weight vector $\mathbf{w}$ that classifies all $\mathbf{a}^k$ as well as possible. This amounts to finding a $\mathbf{w}$ that satisfies as well as possible the following linear system:

$$\begin{array}{l} \mathbf{a}^k\,\mathbf{w} \geq 0 \ \ \forall k \text{ such that } b^k = 1 \\ \mathbf{a}^k\,\mathbf{w} < 0 \ \ \forall k \text{ such that } b^k = -1. \end{array} \tag{3}$$

If the task is linearly separable, the well-known perceptron algorithm [14] is guaranteed to yield in finite time a weight vector that correctly classifies all $\mathbf{a}^k$. The procedure is very simple: start with an arbitrary initial vector $\mathbf{w}_0$, randomly select a sequence of input vectors $\mathbf{a}^k$ and, at each iteration, update the current vector $\mathbf{w}_i$ as follows:

$$\mathbf{w}_{i+1} := \begin{cases} \mathbf{w}_i + \eta_i \ b^k \mathbf{a}^k & \text{if } \mathbf{a}^k \text{ is misclassified by } \mathbf{w}_i \\ \mathbf{w}_i & \text{otherwise,} \end{cases} \tag{4}$$

where η_i is the gain parameter. In fact, when (3) admits a solution, it can be find in polynomial time using an appropriate method for linear programming [15]. In the frequent case where the task is not linearly separable, the best we can ask for is an *optimal* weight vector which correctly classifies as many input vectors as possible. As we have shown in [16, 17], finding an optimal solution turns out to be harder than finding a solution of (3) when it exists. More precisely, the problem is $\mathcal{NP}$-hard and cannot be approximated in polynomial time within every constant factor.
Since the perceptron algorithm never converges for nonlinearly separable tasks, several variants have been proposed and studied for finding optimal weight vectors [8, 9]. In terms of linear systems, these methods aim at finding maximum consistent subsystems of the associated system (3).
The basic idea of the thermal perceptron algorithm [8] is to favor weight updates that aim at correcting errors whose total input $v^k = b^k \mathbf{a}^k \mathbf{w}$ is relatively close to zero. Indeed, since v^k is proportional to the distance from the input to the hyperplane $H = \{\mathbf{x} \in \Re^n | \mathbf{w}\mathbf{x} = 0\}$ defined by the current $\mathbf{w}$, the large weight changes that would be required to correct errors with a large $|v^k|$ are likely to corrupt other well classified $\mathbf{a}^k$. Therefore, the magnitude of the weight modifications decreases exponentially with $|v^k|$. Specifically, the gain parameter in (4) is taken as

$$\eta_i = \frac{t}{t_0} \exp\left(-|v^k|/t\right) \tag{5}$$

where the *temperature* t linearly decreases from an initial value t_0 to 0 in a given number of cycles $C_{\max}$ through the task. This is achieved by setting at the beginning of the cth cycle $t := \gamma t_0$ with $\gamma = 1 - c/C_{\max}$. The annealing process allows to stabilize the weight vector over any given training period.
As shown in [8, 9], the choice of an initial temperature t_0 and of the specific annealing schedule has a strong impact on the quality of the solutions. In particular, faster decreasing schemes provide comparable results more rapidly.

Any variant of the perceptron algorithm can be easily adapted to the optical flow segmentation problem. To cope with noise in the optical flow estimate, the original system (3) of $2n$ equations is replaced by the following system of $4n$ inequalities:

$$\begin{array}{rcl} w_1 + w_2 x_k + w_3 y_k & \geq & d_x(x_k, y_k) - \epsilon \\ -w_1 - w_2 x_k - w_3 y_k & \geq & -d_x(x_k, y_k) - \epsilon \\ w_4 + w_5 x_k + w_6 y_k & \geq & d_y(x_k, y_k) - \epsilon \\ -w_4 - w_5 x_k - w_6 y_k & \geq & -d_y(x_k, y_k) - \epsilon \end{array} \tag{6}$$

where $k \in \{1, 2, ..., n\}$ and $\epsilon > 0$ corresponds to the maximum acceptable error.
As mentioned in section 2, the idea of the approach is to extract from (6) close-to-maximum consistent subsystems iteratively. Therefore, starting with

an arbitrary initial solution $\mathbf{w}_0 \in \Re^6$, an image point (x_k, y_k) is chosen randomly, the corresponding set of inequalities (6) is considered and the error of the current solution $\mathbf{w}_i$ is evaluated with respect to each one of the four inequalities. In this case, the error is defined as the difference between the right and the left hand side terms. For each violated inequality, the current solution $\mathbf{w}_i$ is then updated as in the thermal perceptron:

$$\mathbf{w}_{i+1} := \mathbf{w}_i + \frac{t}{t_0} \exp\left(-|E_i^k|/t\right) \ \mathbf{a}^k \tag{7}$$

where E_i^k denotes the error and $\mathbf{a}^k$ is the coefficient vector of the inequality at hand. The next cycle is started when all image points have been considered in random order.

Eliminating a close-to-maximum consistent subsystem and iterating the procedure until the remaining subsystem is consistent yields a partition of the image into a set of regions corresponding to different motions. Although this simple greedy strategy turns out very effective experimentally, it is not guaranteed to lead in polynomial time to partitions containing a minimum number of consistent subsystems. Indeed, we proved that this minimum cardinality partition problem [18] and the maximum consistent subsystem one [16, 17] are $\mathcal{NP}$-hard as well as hard to approximate. Therefore the best we can do is to devise fast heuristic methods with good average-case behavior.

In fact, it is easily verified that even if maximum consistent subsystems were available at each step, our greedy strategy would not be guaranteed to yield minimum partitions [18]. Nevertheless, the idea of breaking down the overall partition problem into that of determining a sequence of maximum consistent subsystems is particularly attractive in a setting, like image processing, where the ultimate goal is to achieve real-time computation. This is especially true since, in spite of the inherent worst-case complexity of finding maximum consistent subsystems, the thermal variants of the perceptron algorithm exhibit a good experimental behavior [8, 9, 19].

Finally, it is worth noting that, unlike in the usual perceptron applications, the weight vectors determined during the training phase are not used to classify new patterns. Instead, they allow to extract the structure and they provide the parameters of the piecewise linear model of object motion.

4. COMPARISONS WITH OTHER APPROACHES

In order to evaluate the potentialities, properties and limitations of the proposed approach it is worth comparing it with some other ones that have been used till now to tackle the same problem. Only some key points are recalled and a complete analysis has to be omitted for lack of space.

Robust regression analysis is an important statistical tool frequently employed in computer vision for motion analysis [11, 10]. These least square based methods achieve optimum results when the data error distribution is Gaussian. However, they become unreliable if the noise has non-zero mean components and if outliers are present in the data. Outliers are usually considered as large measurements errors or impulse noise corrupting the data. In reality, if we consider the object motion estimation problem the definition of *outliers* is too restrictive since all motion data that do not belong to the *dominant* motion should be considered as outliers. The *breakdown point* of a regression method is the smallest amount of the outlier contamination that may force the value of the estimate outside an arbitrary range. Many robust regression methods have been proposed in the literature [11, 10]. The highest breakdown point of 0.5 is reached by some methods such as the repeated-median and the least median of squares. The complexity of those methods is very high, respectively of the order of $O(n^p \log_p n)$ and $O(n^{(p+1)} \log n)$, where p is the number of parameters of the regression and n is the number of data elements.

In the optical flow segmentation problem, it cannot be guaranteed *a priori* that more that 50% of the data belongs to the same coherent parameter model. Therefore, the data needs to be partitioned before applying robust regression techniques for parameter estimation. Many measures have been proposed for the space partitioning of image features or optical flow data. Statistical confidence measures on the optical flow discontinuities, on the gray level discontinuities, on the position of the foci of expansion [4, 3] have been used to generate hypothesis on object motion structure. The clustering of the data into disjoint or fuzzy sets is then generally based on minimizing *distance* measures among the data and the cluster centroids. However, in most of these methods the number of regions in the partition has to be guessed in advance and all of them have a high degree of complexity. For instance, the k-medoids-like algorithms are characterized by an $O(n^k)$ complexity, where n is the number of data elements and k is predetermined the number of clusters [20].

Among the above-mentioned methods, only the Hough Transform [13] (HT) can, in principle, solve the multiple object motion problem without limitations or heuristic hypothesis on the data structure [12]. But in most applications the HT requires a prohibitive amount of time and space in order to reach a good accuracy.

5. EXPERIMENTAL RESULTS

A number of experiments have been carried out with optical flows generated synthetically and extracted from natural images. Figure 1 reports a synthetic affine optical flow that contains five consistent subsystems of the same dimension. The correct segmentation is shown in figure 2. The parameter values have been generated randomly with some constraints in order to yield realistic values of the flow vectors. Subsystems of the same size have been chosen to test the algorithm in the most critical conditions. Note that no robust regression algorithm can cope with this data unless it is preceeded by a preliminary and, in most cases, problematic clustering stage. Results for subsystems with different dimensions show similar behaviors but faster

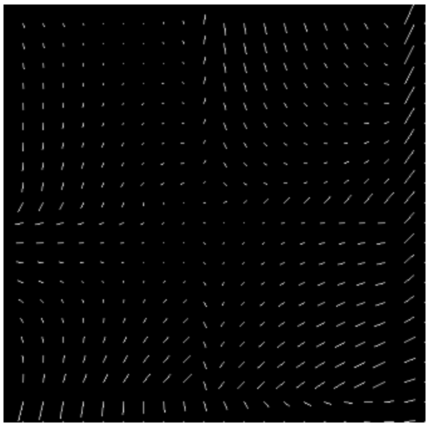

Figure 1: Flow field composed by five synthetically generated affine transformations.

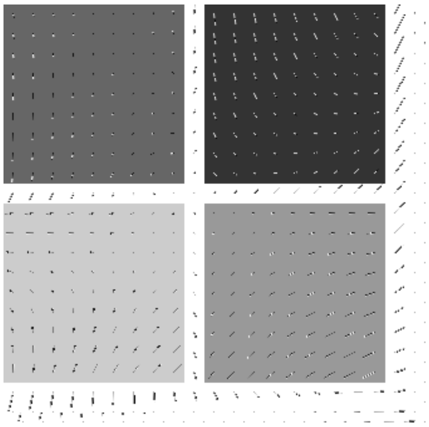

Figure 2: Segmentation of the field into the consistent subsystems. Each gray level corresponds to a consistent subsystem.

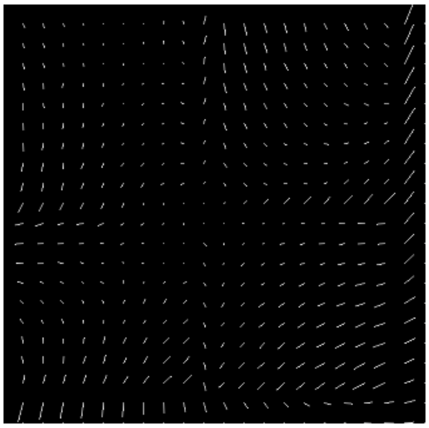

Figure 3: Flow field composed by five synthetically generated affine transformations and superimposed additive gaussian noise.

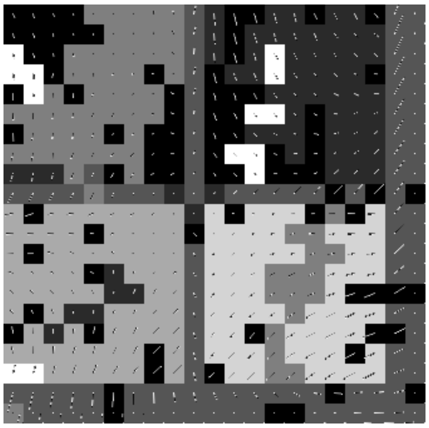

Figure 4: Segmentation of the noisy field into consistent subsystems. Each gray level corresponds to a consistent subsystem.

	w_1	w_2	w_3	w_4	w_5	w_6
Par. subs.1	0.020	0.209	-0.394	-1.828	-0.1255	-0.1165
Estimate	0.0203	0.2094	-0.3938	-1.8277	-0.1254	-0.1166
Noisy est.	0.2543	0.3096	-0.2717	-1.6239	-0.1122	0.0974
Par. subs.2	0.734	-0.473	0.0155	0.29	0.441	0.368
Estimate	0.7342	-0.4729	0.0154	0.2897	0.4409	0.3681
Noisy est.	0.7491	-0.4888	0.0145	0.4718	0.4129	0.3857
Par. subs.3	0.2740	0.283	-0.017	-0.0440	-0.2525	-0.3445
Estimate	0.2739	0.2832	-0.0169	-0.0437	-0.2525	-0.3447
Noisy est.	0.2784	0.2786	-0.0441	0.0015	-0.2576	-0.3363

Table 1: True and estimated parameters for three affine transformations of Figures 2 and 4.

convergence. Figure 3 shows the same flow structure with the addition of pseudo-random gaussian noise with a variance of about 20% of the average data dynamic. The segmentation result is reported in figure 4. Table 1 provides an example of true and estimated parameters in both noisy and non noisy cases. For optical flows without noise, the thermal perceptron algorithm determines the exact segmentation as well as the true motion parameters in less than 300 cycles through all image points. In the case of noisy flows, the relevant structure of the segmentation is detected but some small consistent subsystems appear due to noisy vectors. However, the parameter estimation of the dominant systems remains quite accurate and the spurious subsystems can be easily identified based on their (small) size or poor spatial coherence. In fact, the corresponding noisy vectors could even be corrected by performing an additional optical flow estimation according to the spatial consistency of the current partition.

The results obtained for natural images and presented elsewhere [18], are similar to those for noisy synthetic flows. Indeed, the algorithm can estimate in a very small number of operations (compared to classical approaches), the number of moving objects, their motion parameters and their spatial structure.

Given its computational requirements and performance, our approach fits particularly well into the layered motion representation scheme described in [7] and allows to circumvent the “a priori” clustering problem without facing the structural limitation of classical estimation techniques.

6. CONCLUSION

A new approach based on the extraction of maximum consistent subsystems of linear systems has been proposed for the analysis and parameterization of motion. Simple variants of the perceptron algorithm can be used to tackle the general problem of determining the linear motion parameters of moving objects from image sequences. Unlike in the usual perceptron applications, the goal of the training process is not to generalize to new patterns, but to determine as large as possible consistent subsets of known data (image points) and then to interpret the resulting weight vector components as the motion parameters of the corresponding regions. The experimental results obtained for motion fields generated synthetically and extracted from natural images indicate that our approach allows to overcome some intrinsic limitations of other classical alternatives and compares very favorably in terms of complexity and performance.

References

[1] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performances of optical flow techniques. Technical Report Report no. 299, Dept. of Computer Science University of Western Ontario, London, Ontario, Canada, July 1992.

[2] G. Adiv. Determining three-dimensional motion and structure from optical flow generated by multiple moving objects. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI-7, No. 4, pp. 384–401, July 1985.

[3] A. Rognone, M. Campani, and A. Verri. Identifying multiple motions from optical flow. In G. Sandini editor, editor, *Proceeding of the second European conference on computer vision*, pages 258–266, S. Margherita Ligure, Italy, May 1992.

[4] M. Pertrou, M. Bober, and J. Kittler. Multiresolution motion segmentation. In IEEE Computer Society Press, editor, *Proceedings of the 12th international conference on pattern recognition*, volume 1, pages 743–746, Jerusalem, Israel, October 9-13 1994.

[5] J. Berger, P. Anadan, K. Hanna, and R. Hingorani. Hierarchical model based motion estimation. In G. Sandini editor, editor, *Proceeding of the second European conference on computer vision*, pages 237–252, S. Margherita Ligure, Italy, May 1992.

[6] J. Wang and E. Adelson. Layered representation for motion analysis. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 361–366, New York, USA, June 1993.

[7] S. Hsu, P. Anadan, and S. Peleg. Accurate computation of optical flow by using layered motion representation. In IEEE Computer Society Press, editor, *Proceedings of the 12th international conference on pattern recognition*, volume 1, pages 743–746, Jerusalem, Israel, October 9-13 1994.

[8] M. Frean. A "thermal" perceptron learning rule. *Neural Computation*, Vol. 4, No. 6, pp. 946–957, 1992.

[9] E. Amaldi and C. G. Diderich. On the probabilistic and thermal perceptron training algorithms. Technical Report ORWP-8-94, Department of Mathematics, Swiss Federal Institute of Technology, Lausanne, 1994.

[10] A. M. Leroy P.J. Rousseeuw. *Robust regression and outlier detection.* John Wiley & Sons, Inc, New York, 1987.

[11] P. Meer, D. Mintz, A. Rosenfeld, and D. Kim. Robust regression methods for computer vision: a review. *International journal of computer vision*, Vol. 6, pp. 59–70, 1991.

[12] R.A. Samy and C.A. Bozzo. Moving object recognition using motion enhanced Hough transform. In V. Cappellini and A.G. Costantinides, editors, *Digital Signal Processing-84*, pages 770–775, Amsterdam, The netherlands, 1984.

[13] J. Illingworth and J. Kittler. The adaptive hough transform. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI-9, No. 5, pp. 690–698, September 1987.

[14] M. L. Minsky and S. Papert. *Perceptrons: An introduction to computational Geometry.* MIT Press, Cambridge, MA, 1988. Expanded edition.

[15] N. Karmarkar. A new polynomial time algorithm for linear programming. *Combinatorica*, Vol. 4, pp. 373–395, 1984.

[16] E. Amaldi. On the complexity of training perceptrons. In T. Kohonen, K. Mäkisara, O. Simula, and J. Kangas, editors, *Artificial Neural Networks*, pages 55–60, Amsterdam, 1991. Elsevier science publishers B.V.

[17] E. Amaldi and V. Kann. The complexity and approximability of finding maximum feasible subsystems of linear relations. Technical Report ORWP-11-93, Department of Mathematics, Swiss Federal Institute of Technology, 1993. To appear in Theoretical Computer Science.

[18] M. Mattavelli and E. Amaldi. Manuscript in preparation, 1995.

[19] E. Amaldi. *From finding maximum feasible subsystems of linear systems to feedforward neural network design.* PhD thesis, Department of Mathematics, Swiss Federal Institute of Technology, Lausanne, 1994.

[20] P.J. Rousseeuw L. Kaufman. *Finding groups in data.* John Wiley & Sons, Inc, New York, 1990.