A combinatorial optimization approach to extract piecewise linear structure from nonlinear data and an application to optical flow segmentation

Edoardo Amaldi§* and Marco Mattavelli††

§School of Operations Research and Theory Center

Cornell University, Ithaca, NY 14853, USA

amaldi@cs.cornell.edu

†Signal Processing Laboratory, Swiss Federal Institute of Technology,

CH-1015 Lausanne, Switzerland

mattavelli@de.epfl.ch

December 9, 1997

[1] Manuscript received:

*E.A. was partially funded by a Postdoctoral fellowship of the Swiss National Science Foundation.

†M.M. is currently with the Integrated System Center (C3I), Swiss Federal Institute of Technology, CH-1015 Lausanne, Switzerland.

Abstract

Many tasks arising in the analysis of visual information have been considered as ill-posed linear inverse problems. Due to the importance of discontinuities, piecewise linear models are often appropriate. However, although there have been several attempts to extend regularization theory to deal with discontinuities, simultaneously finding the piecewise decomposition and the corresponding submodel parameters is a challenging problem. In this paper, a new combinatorial optimization approach based on partitioning inconsistent linear systems into a minimum number of consistent subsystems (MIN PCS) is proposed. We show that, in spite of the worst-case complexity of MIN PCS, satisfactory approximate solutions can be obtained using simple variants of a classical relaxation procedure for solving systems of linear inequalities. The potential of our approach is demonstrated on the difficult problem of detecting and parameterizing the 2-D motion of multiple moving objects from a sequence of images, in particular on the problem of optical flow segmentation. Experimental results obtained for synthetic and natural images are reported and the advantages with respect to alternative methods are discussed.

1 Introduction

The interpretation of visual information consists of extracting high-level features such as shapes, objects, 3-D structure and motion from image sequences, namely, from the raw data (pixel values). Since these tasks involve recovering properties of 3-D scenes from 2-D images, they can be regarded as ill-posed inverse optics problems [1]. In particular, the solution may not exist, may not be unique or may not depend continuously on the data due to the ambiguities arising from the 2-D projection and the noise or errors in the data.

Regularization theory (see [1]) has been proposed as a unified and standard way of finding approximate solutions to ill-posed problems. The main idea is to introduce *a priori* knowledge by using variational principles that impose constraints on the admissible solutions or by making statistical assumptions on the solution space (see [2] and the included references). The global *smoothness* constraint is a typical example of widely used a priori assumption. A variety of computational vision problems such as, for instance, edge detection, surface reconstruction and optical flow estimation, have first been tackled using quadratic functionals and Tikhonov-like stabilizers [1, 3]. Since discontinuities frequently convey crucial information, classes of nonlinear and spatially non-invariant regularization methods [4] as well as a probabilistic framework based on Markov random fields (MRF) [2] have been proposed. The aim is to deal with more realistic piecewise smoothness constraints. However, the introduction of discontinuities within this type of approaches (using explicit line processes or assuming prior distributions) turns out to be a delicate issue. Indeed, line processes are not suited to deal with occlusion because they do not allow propagation of information between occluded regions and probabilistic regularization require a priori knowledge on the distributions specifying local dependencies. Moreover, the resulting methods are computationally intensive and therefore not appropriate when the ultimate goal is to achieve efficient and/or real-time processing.

In several image processing problems such as motion analysis, the goal is to extract the structure from the data (e.g., detect discontinuities) as well as to estimate the parameters of the model that represents the feature or property under consideration. Piecewise linear models are often attractive since they enable approximation of complex nonlinear phenomena but are still simple enough due to local linearity. Unfortunately, piecewise decomposition and parameter estimation appears as a difficult problem. Conventional approaches in which one first partitions the data using some clustering methods [5] and then estimates the parameters associated with each submodel using robust regression techniques [6] have major drawbacks. In particular, the number of linear submodels needs to be guessed in advance and, even more importantly, clustering methods do not take into account the type of submodel used. In principle, the Hough Transform (HT) and its variants [7, 8] can solve a wide range of model identification problems without requiring any a priori assumption. However, their computational requirements to guarantee a reasonable accuracy can be prohibitive even for small problems.

In this paper we present a combinatorial optimization approach to simultaneously determine the structure and estimate the parameters of piecewise linear models. In Section 2 we show that this problem can be formulated as that of Partitioning linear systems into a MINimum number of Consistent Subsystems. The computational complexity of this new combinatorial optimization problem, referred to as MIN PCS, is discussed in Section 3. In Section 4 we describe a simple greedy strategy which provides good approximate solutions in a short amount of time. Section 5 compares our approach with conventional alternatives such as robust regression, clustering methods and the Hough transform. In Section 6 the potential of our approach is demonstrated on the challenging problem of detecting and parameterizing the 2-D motion of multiple moving objects in image sequences. In particular, we focus here on the segmentation of the optical flow. 2-D motion estimation has attracted a considerable attention because it is an intermediate step towards 3-D motion estimation [9] and it is useful in a number of settings such as, for instance, object tracking, robot guidance and in new video compression standards like MPEG-4 (see [10] and included references). Finally, concluding remarks are mentioned in Section 7.

2 Piecewise linear model estimation and MIN PCS problem

As is well known, simple linear models lead to formulations in terms of linear systems such as:

$$A\,\mathbf{x} = \mathbf{b} \tag{1}$$

where A is a $p \times n$ matrix and $\mathbf{b}$ a p-dimensional vector. In the presence of discontinuities in the high-level feature or property under consideration, (1) is in general inconsistent, i.e., no solution satisfies all linear equations. In other words, one is confronted with over-determined systems (p is typically much larger than n) for which classical least-mean-square error solutions are not appropriate. Also robust regression techniques [6], which assume that the underlying linear estimation problem admits a solution and that (1) is inconsistent because of spurious or noisy data, have a limited range of application. Indeed, in many situations inconsistency is not actually due to the presence of outliers but to the fact that a linear model is too simple to model the actual complexity of the problem at hand. This is especially true for problems involving discontinuities. Piecewise linear models often provide a good trade-off between representation power and local simplicity, but the issue is then to determine the piecewise decomposition as well as to estimate the parameters of each linear submodel.

In this paper we consider a combinatorial optimization problem which allows formalization of the fundamental problem of piecewise linear model estimation.

> MIN PCS: Given a possibly inconsistent linear system $A\mathbf{x} = \mathbf{b}$ with $A \in \mathbf{R}^{p \times n}$ and $\mathbf{b} \in \mathbf{R}^p$, find a Partition of this system into a MINimum number of Consistent Subsystems.

According to the well-known Occam razor principle [11], we look for the "simplest" piecewise linear model consistent with the data, which is most likely to be the correct one. Here complexity is measured in terms of the number of linear submodels.

The MIN PCS formulation is very attractive because it provides a natural way of addressing simultaneously the two fundamental issues: piecewise decomposition and parameter estimation. Given any solution of MIN PCS the partition indicates the piecewise decomposition and a solution of each consistent subsystem provides a set of parameter values for the corresponding submodel.

From the point of view of ill-posed linear inverse problems, this amounts to look for the smallest number of smooth (linear) submodels that are needed to account for the data. In other words, it enables one to enforce piecewise smoothness constraints on the solutions and to deal with discontinuities without making any particular assumption on their geometry[1]. This is in contrast with probabilistic

[1] The reader is referred to [10] for a more detailed discussion of discrete linear inverse problems with discontinuities and piecewise linear models.

regularization approaches where piecewise smooth random fields are obtained by imposing a set of constraints on their local properties and, in particular, on the shape of the discontinuities [2]. Such local constraints try, for instance, to favor discontinuities arranged in piecewise smooth lines and to prevent the formation of smooth patches (regions of the image corresponding to the same submodel) that are too small. Clearly, since only neighboring sites of Markov random fields interact, only local dependencies (features) can be taken into account.

A unique aspect of our MIN PCS formulation is that it aims at minimizing the number of submodels. The choice of this objective function provides a natural way of avoiding too small patches that are meaningless, and of extracting nonlocal dependencies (nonlocal features) since the equations in (1) that can be accounted for by the same submodel are grouped together regardless of the location between the corresponding image pixels. It has to be emphasized that this is achieved without making any a priori assumption on the feature under consideration and on the type of piecewise linear model. In particular, the model is not restricted to be a continuous and single-valued function. Clearly, the type and the quality of the resulting model depends on how representative is the data at hand.

In practice, to cope with noisy data, a noise tolerance threshold $\varepsilon > 0$ is selected and each equation $\mathbf{a}^k\mathbf{x} = b_k$, where $\mathbf{a}^k$ denotes the kth row of A and b_k the kth component of $\mathbf{b}$ with $1 \leq k \leq p$, is replaced with the two complementary inequalities:

$$\mathbf{a}^k\mathbf{x} \leq b_k + \varepsilon \qquad \mathbf{a}^k\mathbf{x} \geq b_k - \varepsilon \tag{2}$$

that have to be simultaneously satisfied. If equations of the original system are expected to be affected by different noise levels, different thresholds can of course be used.

It is worth noting that these variants of MIN PCS have a simple geometric interpretation. As shown in Figure 1, if $b_k = 0$ for all $1 \leq k \leq p$ they amount to finding the minimum number of hyperslabs of thickness 2ε (obtained by moving a hyperplane through the origin apart by ε in the opposite directions) such that all the points $\mathbf{a}^k \in \mathbf{R}^n$ corresponding to the coefficient vectors are contained in at least one hyperslab. The normal vectors defining these hyperplanes clearly correspond to the solutions of the consistent subsystems in a minimum partition. In the general case with arbitrary right hand sides b_k, the hyperplanes are not required to go through the origin and the geometric interpretation is slightly more involved.

3 Worst-case complexity of MIN PCS

Since to the best of our knowledge MIN PCS has not yet been discussed in the literature, the question of its inherent computational complexity arises. If the points $\mathbf{a}^k$ are in general position[2], the problem is trivial since any subsystem with up to n equations is consistent and all larger subsystems are inconsistent. However, can every instance of MIN PCS be solved in a number of operations that grows polynomially with its size (n, p and the magnitude of the coefficients) or is it at least as hard to solve in the worst-case as many important problems such as the famous traveling salesman problem [12]? The theory of NP-completeness [13, 14] provides a framework to address this question.

Theorem 1 MIN PCS *is* NP*-hard, i.e., it cannot be solved in polynomial time unless* P = NP*.*

The proof is given in the Appendix. In fact, we show that it is NP-complete to decide whether inconsistent systems can be partitioned into 2 consistent subsystems. Thus, assuming the fundamental conjecture of complexity theory —P $\neq$ NP— is true, all instances of MIN PCS cannot be solved to optimality in polynomial time.

In most applications, however, optimum solutions are not required and approximate solutions that can be found in a reasonable amount of time are of great practical value. Now our proof also implies that it is NP-hard to find near-optimal solutions. Indeed, if there exists a polynomial time algorithm that is guaranteed to yield for every instance a partition with a number of consistent subsystems which exceeds the minimum one by a multiplicative factor $\rho < 3/2$, then it would yield for any bipartitionable instance a solution of size strictly less than $3/2 \cdot 2 = 3$, that is, with 2 consistent subsystems. But, since such an algorithm would solve an NP-complete problem in polynomial time, it follows that MIN PCS cannot be approximated within a factor $\rho < 3/2$, unless P = NP.

Although this result provides important insight into the inherent difficulty of MIN PCS, we must be aware that it is based on a worst-case analysis and that, therefore, it does not rule out the existence of efficient heuristics with a good average-case behavior. Experience with other difficult combinatorial optimization problems has indeed shown that the computational complexity predicted by worst-case results need not correspond to that of common real-world instances.

It is noteworthy that MIN PCS is trivial for inconsistent systems of inequalities since any such system can be partitioned into two consistent subsystems. This fact is obvious for homogeneous systems and is easily verified for inhomogeneous ones [15]. But it is not true if the system of inequalities derives from a

[2] A set of points in $\mathbf{R}^n$ is in general position if no subset of $k+1$ points is contained into an affine space of dimension $k-1$.

system of equations like (2) and one looks for solutions that satisfy pairs of complementary inequalities associated with the same original equation. The polynomial time reduction given in the Appendix can be readily extended to this case.

Thus, quite intuitively, designing piecewise linear models turns out to be more difficult than estimating simple linear ones. Indeed, MIN PCS is NP-hard while standard least-mean-square solutions can be found in polynomial time.

4 A greedy algorithm

Since in practical applications we are interested in finding approximate solutions of MIN PCS rapidly, we adopt a simple greedy approach and subdivide the overall partition problem into a sequence of subproblems. This type of strategy has been extensively used in the design of fast heuristics for a number of NP-hard optimization problems.

Consider the inconsistent system $A\mathbf{x} = \mathbf{b}$ corresponding to a single linear model that is not powerful enough to model the underlying phenomenon. The idea is to find a close-to-minimum partition into consistent subsystems by iteratively extracting maximum consistent subsystems, i.e., consistent subsystems containing a maximum number of equations. Clearly, starting with the original system, eliminating a close-to-maximum consistent subsystem at each step and iterating the process until the remaining subsystem is consistent yields a partition into consistent subsystems. The strategy can also be applied to the MIN PCS version with pairs of complementary inequalities (2).

As we shall see, this greedy strategy is well motivated from the application point of view and is effective experimentally. However, according to Theorem 1, it is not guaranteed to yield approximate solutions that are within a factor $\rho < 3/2$ of the optimum.

Before discussing some theoretical and practical properties of the overall greedy procedure, we focus on the subproblem that has to be dealt with repeatedly.

4.1 Extraction of close-to-maximum consistent subsystems

At each step of the greedy strategy, one looks for a maximum consistent subsystem of a linear system. More precisely, given a system $A\mathbf{x} = \mathbf{b}$ and a tolerance threshold $\varepsilon > 0$, one seeks an $\mathbf{x} \in \mathbf{R}^n$ that satisfies as many pairs of complementary inequalities $\mathbf{a}^k\mathbf{x} \leq b_k + \varepsilon$ and $\mathbf{a}^k\mathbf{x} \geq b_k - \varepsilon$ as possible.

Several variants of the general problem of finding MAXimum Consistent Subsystems of linear systems, referred to as MAX CS, arise in other fields such as machine learning and operations research (see [16] and the included references). In [17] we investigated the computational complexity of MAX CS and in [16, 18] some variants of the so-called perceptron method [19] studied in the machine learning literature. Although MAX CS is NP-hard to approximate within some constant factors [17], those thermal[3] perceptron variants [20, 16, 18] provide good solutions in a short amount of time.

In this subsection we show that a thermal version of the Agmon-Motzkin-Schoenberg relaxation method (AMS) for solving consistent systems of linear inequalities [21, 22, 23] can be used to find large consistent subsystems of (2).

The AMS method is a simple iterative procedure that generates a sequence of estimates. Given a consistent system $A\mathbf{x} \leq \mathbf{b}$, start with an arbitrary initial guess $\mathbf{x}_0 \in \mathbf{R}^n$ and consider each inequality cyclically but in a random order. Suppose that at iteration i the kth inequality $\mathbf{a}^k\mathbf{x}_i \leq b_k$ is considered. Then the current estimate $\mathbf{x}_i$ is updated as follows:

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \delta_i \mathbf{a}^k \quad \text{where} \quad \delta_i = \max\{0, \lambda_i(\mathbf{a}^k\mathbf{x}_i - b_k)\} \tag{3}$$

and the parameter λ_i satisfies $0 < \lambda_i < 2$. In other words, at each iteration all but one inequalities are relaxed and the current estimate is updated only if it violates the inequality at hand.

Geometrically, the current estimate $\mathbf{x}_i$ can be viewed as a point in $\mathbf{R}^n$ and the kth inequality defines a hyperplane $H_k = \{\mathbf{x} \in \mathbf{R}^n | a^k\mathbf{x} = b_k\}$. If $\mathbf{x}_i$ lies on the negative side of H_k (kth inequality is satisfied) then it is unchanged. Otherwise, $\mathbf{x}_{i+1}$ is obtained by making a step along the line through $\mathbf{x}_i$ and perpendicular to H_k. For $\lambda_i = 1$, $\mathbf{x}_{i+1}$ is the orthogonal projection of $\mathbf{x}_i$ onto H_k, see Figure 2. For other values, $\mathbf{x}_{i+1}$ lies on the line connecting $\mathbf{x}_i$ and its orthogonal reflexion with respect to H_k.

For any consistent system, this method is guaranteed to yield a solution in a finite but in the worst-case exponential number of iterations [24]. Although more sophisticated stopping criteria have been proposed [24], in practice one usually stops after a predefined maximum number of cycles C through the p inequalities.

We now describe a thermal variant of the AMS relaxation method which provides close-to-maximum consistent subsystems of (2) in a short amount of time. The problem when dealing with inconsistent systems is that the standard procedure never terminates (violated inequalities always trigger updates) and in a worst-case scenario it can even lead in a single iteration from an estimate that satisfies a maximum number of inequalities to a worst possible one. To make the procedure well behaved for inconsistent systems, the basic idea is to favor updates of the current estimate $\mathbf{x}_i$ which aim at correcting

[3]The adjective "thermal" refers to the fact that a temperature parameter is used like in simulated annealing.

unsatisfied inequalities with a relatively small *violation*

$$v_i^k = \begin{cases} \mathbf{a}^k\mathbf{x}_i - b_k - \varepsilon & \text{if } \mathbf{a}^k\mathbf{x}_i \geq b_k + \varepsilon \\ b_k - \varepsilon - \mathbf{a}^k\mathbf{x}_i & \text{if } \mathbf{a}^k\mathbf{x}_i \leq b_k - \varepsilon \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Indeed, the large modifications that would be required to correct unsatisfied inequalities with large v_i^k are likely to corrupt other inequalities that $\mathbf{x}_i$ satisfies. As in [20], it is appropriate to pay decreasing attention to unsatisfied inequalities with large violations, namely, to perform at the beginning all updates prescribed by the standard procedure and then only those which aim at correcting unsatisfied inequalities with progressively smaller and smaller v_i^k. This can be achieved by letting the magnitude of the updates δ_i decrease exponentially with v_i^k and by introducing a control parameter T with which the v_i^k are compared. The algorithm is described below. T is the so-called *temperature* parameter and it is linearly decreased from an initial value T_0 to 0 in the predefined maximum number of cycles C. Specifically, at the beginning of the c^{th} cycle $T = \gamma T_0$ with $\gamma = 1 - \frac{c}{C}$. See the Appendix for more details about this annealing schedule, including an appropriate choice of T_0.

Algorithm 2

Initialization: Pick arbitrary $\mathbf{x}_0 \in \mathbf{R}^n$, set cycle counter $c = 1$, select predefined number of cycles C and choose initial T_0

while $c \leq C$ **do**

  Initialize set of indices $I = \{1, \ldots, p\}$

  repeat —cycle through all pairs of inequalities—

    Pick index k_i equiprobably and without replacement from I

    Compute violation $v_i^{k_i}$ for the selected pair of inequalities

    Set $T = (1 - \frac{c}{C}) \cdot T_0$ and $\delta_i = \frac{T}{T_0} \exp\left(\frac{-v_i^{k_i}}{T}\right)$

    if $\mathbf{a}^{k_i}\mathbf{x}_i \geq b_{k_i} + \varepsilon$ **then** $\mathbf{x}_{i+1} = \mathbf{x}_i - \delta_i \, \mathbf{a}^{k_i}$

    else if $\mathbf{a}^{k_i}\mathbf{x}_i \leq b_{k_i} - \varepsilon$ **then** $\mathbf{x}_{i+1} = \mathbf{x}_i + \delta_i \, \mathbf{a}^{k_i}$

      else $\mathbf{x}_{i+1} = \mathbf{x}_i$

    Set $I = I - \{k_i\}$

  until $I = \emptyset$

  Update T_0 (cf. Appendix) and set $c = c + 1$

end

return $\mathbf{x}_{p\cdot C}$

Note that, unlike in regression methods, inequalities with a large violation with respect to the current temperature yield small modifications of the current estimate and those that are satisfied do not yield any modification. For very high values of T, any violated inequality yields a significant update, while for low values, only those with small violations with respect to T yield significant updates.
Instead of returning the last estimate $\mathbf{x}_{p\cdot C}$ as the approximate solution, one can return the best estimate generated during the C cycles. Therefore it suffices to check whether each new estimate $\mathbf{x}_{i+1}$ satisfies a larger number of pairs of inequalities (2) than the best one generated so far. See [18] for the asymptotic convergence properties of closely related variants of the thermal perceptron procedure.

Finally, it is worth noting that implicit enumeration methods, like the general technique for maximizing functions of linear relations presented in [25], can be adapted to find optimal solutions in a finite but in the worst-case exponential amount of time. However, they are only applicable to very small instances.

4.2 Properties of the greedy strategy

The size of the partitions obtained with the above algorithm depends on the global strategy as well as on the effectiveness of the method used to deal with the subproblems. Given the greedy approach, it is in fact neither necessary nor desirable to look at each step for a largest consistent subsystem. In the Appendix we show by means of a simple example that even if maximum consistent subsystems are available at each step, the greedy strategy is not guaranteed to yield minimum partitions. According to this example, the resulting partitions may depend on the order in which the maximum consistent subsystems are extracted. More precisely, some equations may be assigned to different consistent subsystems depending on the sequence of extractions. Since this kind of ambiguity is inherent to the data (some data points may be consistent with several submodels), it can be resolved by taking into account the physical nature of the problem (e.g., the location of the corresponding pixels in the image).

It is worth emphasizing that the idea of breaking down the original problem into that of determining a sequence of maximum consistent subsystems is particularly attractive in image processing settings in which the ultimate goal is to achieve on-line processing. First, such greedy algorithms have low computational requirements that in many applications may allow real-time implementations. Second, it is appropriate to look first for the dominant submodels, i.e., those which account for the largest fraction of the data points, and then focus on smaller and smaller submodels. This amounts to considering progressively higher and higher resolutions.

Note that in practical problems the size of the consistent subsystems that are extracted is typically much larger than the number of variables n. In the presence of noisy or spurious data, once the actual structure has been extracted the points that are left are usually in general position and they can be partitioned into consistent subsystems of size at most n. Such consistent subsystems can be clearly discarded since they do not contain any relevant modeling information. Thus, from an application point of view, the exact minimum partition is not required and a greedy heuristic which looks first for the largest consistent subsystems is very appropriate.

Finally, to achieve higher robustness for noisy data, a least-mean-square solution can be determined for each subsystem of the partition provided by our algorithm. Thus, piecewise linearity is taken into account in the partition process and robustness is guaranteed in submodel parameter estimation.

5 Comparison with some alternative approaches

Before presenting an application of our MIN PCS-based approach to 2-D motion analysis, we briefly compare it with three conventional alternatives.

5.1 Robust regression

Over-determined linear systems are frequently tackled using regression methods. Robust regression techniques [6] have been devised to deal with outliers and have been applied to a number of signal and image processing problems [26]. The *breakdown point* of such a method is the smallest fraction of outliers contaminating the data that can make the estimation error unbounded. Among the robust estimators that have been proposed in the literature, the repeated-median and the least-median-of-squares ones have the highest breakdown point of 0.5 [6, 26]. Unfortunately, they have very high computational requirements, namely an $O(p^n \log_n p)$ and respectively an $O(p^{(n+1)} \log p)$ time complexity, where p is the number of data points and n is the number of parameters to be estimated. Variants of the least-median-of-squares method using Monte Carlo-like techniques have a lower complexity but they may fail with some nonzero (even though low) probability [26].

In practice, when the discontinuities are not known a priori, the definition of outliers is too restrictive. Since all the data that does not belong to the dominant subsystem must be considered as outliers, robust regression techniques cannot be used in the frequent situations in which more than 50% of the data is not accounted for by the same linear submodel. Note that our MIN PCS-based algorithm does not share this fundamental limitation and has a much lower time complexity.

5.2 Clustering methods

To try to circumvent the breakdown point limitation, one may decompose the problem into two distinct phases. First, one partitions the data into a certain number of components using clustering methods [5] and then one estimates each linear submodel using robust regression techniques. The data is usually subdivided into disjoint or fuzzy sets by minimizing some distance measure between the data points and the cluster centroids. Not surprisingly, the number of clusters, which is usually unknown and needs to be guessed in advance, has a strong impact on the results. But, since even the simplest clustering problems are NP-hard [12], the high computational requirements rule out a trial-and-error procedure. If the number of candidate centroids is arbitrarily limited to a certain value c using some thresholding criteria, the time complexity is still $O(p\frac{c!}{(c-k)!\cdot k!})$, where k is the number of clusters, p is as above and obviously $c > k$ [5]. An even more important limitation of such two-phase approaches is that clustering methods may yield meaningless partitions in terms of piecewise linear models since they do not take into account the type of submodels (linear) used. An illustrative example arising in our 2-D motion analysis application will be presented in Section 6.3. Clearly, the MIN PCS-based algorithm does not have these serious shortcomings.

5.3 Hough transform

The Hough Transform is a general class of techniques for linear and nonlinear model identification [7, 8]. Local information is used to accumulate evidence for some particular sets of parameter values of the model under consideration. Being relatively insensitive to noise and partially incorrect data, HT has been proposed, for instance, to determine shape, motion and geometric transformation parameters [27, 28, 29]. In the classical HT, the first step is to map each feature in the data space onto a multidimensional surface in the parameter space corresponding to all combinations of the parameter values consistent with it (e.g., hyperplanes for linear models). By selecting some reasonable range for each parameter, attention is then restricted to a limited region that is subdivided into hypercells. The second step amounts to identifying a hypercell that is hit by the largest number of surfaces, i.e., a set of parameter values that account for the largest fraction of the data. The search for such peaks is performed using exhaustive (see Figure 3) or reduced strategies. The accuracy of the parameter estimation clearly depends on the quantization, i.e., the size of the hypercells.

Unlike robust regression and clustering, the HT enables, in principle, the solution of a wide range of model identification problem and, hence, in particular the estimation of piecewise linear models without a priori assumptions on the data structure. In practice, however, the high dimensionality and the fine resolution needed to guarantee a reasonable accuracy require very large amounts of memory and computation time even for small problems, see for instance [9]. In some cases, intelligent iterative and adaptive coarse to fine search strategies [28, 29] can be applied. In others, deterministically or randomly selected subsets of the data are used in order to make some hypotheses or eliminate outliers in conjunction with other heuristics [9, 30]. However, even if the storage requirements are reduced, the order of complexity of the HT remains very high for models with more than a few parameters. Moreover, variants with refined search strategies (see [7, 8]) that in some cases considerably reduce the computational requirements may not be appropriate when more than a single model have to be simultaneously identified (e.g., the submodels of a piecewise linear one). In the Randomized HT [30], for instance, one may have to sample a much larger number of subsets of data points in order to achieve the ad-hoc threshold indicating the presence of a probable peak. In coarse to fine strategies such as the adaptive HT [28], the initial stages where a coarse quantization is considered may fail to detect some of the peaks corresponding to the submodels to be identified. Indeed, relatively small peaks may be hidden in a noisy background.

Although our MIN PCS-based algorithm and the HT applied to piecewise linear model estimation present some similarities, there are some important differences, see Figure 3 b). Unlike regression methods, both of them look for parameter values consistent with the largest fraction of data points. But, while the cost of the HT grows with the number of hypercells (quantization) and of parameters n, the extraction of each close-to-maximum consistent subsystem involves scanning C times the pairs of inequalities (2) and updating the current estimate accordingly. Hence, each step of our greedy algorithm has an $O(n \cdot p \cdot C)$ complexity and no additional storage is required for higher accuracy. An increase in the desired level of accuracy only increases the total number of thermal AMS iterations since the final partition may contain a larger number of consistent subsystems and each one of them may require a larger number of AMS updates to find a near-optimal solution. However, since the number of pairs of inequalities decreases after the extraction of each consistent subsystem, the computational requirements involved at each step tend to decrease. Thus, our search strategy presents considerable advantages from the computational point of view.

6 An application to 2-D motion analysis

We now briefly present an example of application of our MIN PCS-based approach to the important and classical problem of detecting and parameterizing the 2-D motion of multiple moving objects from image sequences. As mentioned at the end of Section 1, 2-D motion estimation represents a fundamental step towards a global dynamic scene understanding and can be helpful, among others, for describing video sequences in terms of a small number of coherently moving components. In spite of the considerable research effort, estimating motion from image sequences containing multiple moving objects remains a challenging problem. As discussed in [10], the methods proposed in the literature may be subdivided into two wide classes: those based on the optical flow, i.e., the 2-D field of instantaneous velocities of brightness patterns, and those based on motion models. While the former methods involve first the computation of the optical flow and then its interpretation [9, 31, 32, 33, 34, 35, 36], the latter ones, known as parametric methods, try to extract the motion parameters directly from the image intensities [37, 38, 39, 40, 41, 42, 43]. Parametric methods often lead to a layered representation in which each layer should correspond to a different moving component.

6.1 Previous and other work

All existing methods for motion estimation and segmentation, face some kind of serious difficulties when the scene contains multiple moving components instead of a single or dominant one. The central issue is to deal with discontinuities without knowing them a priori.

Although it has been generally suggested that the motion parameters may be extracted by segmenting the optical flow and then applying linear regression, there are some important loopholes. First, the problem of optical flow segmentation proves to be a *chicken and egg* problem. Indeed, it would be easy to solve knowing either the location of the moving objects or their motion parameters. Second, given the presence of discontinuities, the error in the optical flow estimation is usually non-Gaussian and linear regression methods do not work well. However, frameworks in which outliers are detected and rejected do not allow simultaneous estimation of different motion components.

Several algorithms that identify multiple motions from pre-computed optical flow fields rely on clustering techniques [32, 40]. As previously mentioned, these algorithms have two main drawbacks. First, the number of moving components has to be estimated. Second, working only with optical flow vector components and neglecting the dependence of those amplitudes on the reference coordinates, may not suffice to recover the actual motion components. Not surprisingly, clustering-based approaches such as in [32, 40] present a high degree of arbitrariness.

Model-based methods face the same type of fundamental difficulty [40, 43]. The major issue is to partition the images into layers corresponding to different coherently moving components. In some approaches, one tries to identify and extract dominant components iteratively. But the existence of a dominant component at each step is, of course, a very restrictive assumption that is not satisfied by most generic image sequences. Note that the threshold on the percentage of the data needed to be considered as dominant (e.g., at least 50% for robust regression techniques) imposes further restrictions.

A class of methods based on *mixture models* has recently been investigated (see for instance [42, 43, 36, 44]). Although such models can be fit to the image data or to the optical flow measurements, they rely on the assumption that the input data is generated by a given number of smooth motion groups and the algorithms involved (variants of the EM optimization method and MRF computations or mean field approximations) have high computational requirements. Moreover, the number of submodels in the mixture (or at least an upper bound on it) needs to be guessed in advance [42, 43]. Indeed, the automatic way of estimating the number of submodels mentioned in [36, 44] refers to the fact that, if the algorithm is initialized with a large enough number of submodels, then redundant submodels will have the same parameter values. Clearly, selecting a reasonable upper bound is a delicate problem. Larger bounds are safer but they increase the computational requirements further. In addition, the above-mentioned property is only valid if the distribution of the residuals between the observed measurements and the ones predicted according to the various submodels are Gaussians with a given variance σ. Not surprisingly, the number of distinct submodels in the mixture depends on σ.

Markov random fields [2, 45], which take into account discontinuities, are also confronted with the problem of estimating the number of submodels. Moreover, extensions of MRFs to 2-D piecewise smooth models implicitly require that regions, assigned to the same submodel, be contiguous since they only integrate local information [45]. Therefore these approaches face serious difficulties when dealing with occlusions which frequently occur in natural sequences and which give rise to disconnected regions corresponding to the same motion components.

6.2 MIN PCS formulation of optical flow segmentation

We now describe how the optical flow segmentation problem can be easily formulated in terms of MIN PCS[4]. Let us consider an image sequence that evolves in time with the apparent motion given

[4] A preliminary version of the formulation was presented in [46].

by equation:

$$I(x,y,t) = I(x - V_x(x,y), y - V_y(x,y), t-1), \tag{5}$$

where x and y are the image coordinates, $I(\cdot)$ is the image intensity and $V_x(x,y)$ and $V_y(x,y)$ denote the image displacements at each point. Under reasonable assumptions (see for instance [9]), the 2-D apparent motion of an object can be approximated by an affine transformation given, for every image point (x,y), by:

$$\begin{aligned} V_x(x,y) &= w_1 + w_2 x + w_3 y \\ V_y(x,y) &= w_4 + w_5 x + w_6 y \end{aligned} \tag{6}$$

where $w_1, \ldots, w_6$ are the parameters.

Let us assume that the optical flow has been estimated using a standard technique [31], i.e., we have a vector $(V_x(x_k,y_k), V_y(x_k,y_k))$ for p image points (x_k,y_k) with $1 \le k \le p$.

In the presence of multiple motion components, the overall system consisting of the p pairs of equations (6) in the six variables $w_1, \ldots, w_6$ is inconsistent. Since for accurate optical flow estimates the regions with the same affine motion correspond to consistent subsystems, the optical flow segmentation problem amounts to partitioning the above inconsistent system into r consistent subsystems containing pairs of equations (6) and of finding a set of corresponding solutions $w_1^l, \ldots, w_6^l$ with $1 \le l \le r$.

To cope with noise in the optical flow estimate, each pair of equations (6) is replaced by the following group of four inequalities:

$$\begin{aligned} w_1 + w_2 x_k + w_3 y_k &\ge V_x(x_k,y_k) - \varepsilon \\ -w_1 - w_2 x_k - w_3 y_k &\ge -V_x(x_k,y_k) - \varepsilon \\ w_4 + w_5 x_k + w_6 y_k &\ge V_y(x_k,y_k) - \varepsilon \\ -w_4 - w_5 x_k - w_6 y_k &\ge -V_y(x_k,y_k) - \varepsilon \end{aligned} \tag{7}$$

where $1 \le k \le p$ and $\varepsilon > 0$ is the maximum tolerable error. To partition the image in as few regions as possible, we consider the generalization of the MIN PCS problem to systems with such groups of four inequalities. The thermal AMS algorithm can be immediately extended to this case in which we look at each step for consistent subsystems containing as many groups (7) as possible. Note that here x_k, y_k with $1 \le k \le p$, are the coefficients and $w_1, \ldots, w_6$ are the unknowns.

The above formulation and the proposed algorithm provide a natural way of addressing the problem of optical flow segmentation and thereby that of detecting and parameterizing 2-D multiple object motion. From the application point of view, it is a good strategy to look iteratively for maximum consistent subsystems since all points that can be accounted for by the same affine motion parameter values should be associated to the same motion component. It is worth emphasizing that our MIN PCS-based approach does not require any dominant motion or any a priori knowledge, e.g., no assumptions on the number, on the spatial location and on the connectivity of the motion components. Moreover, it has much lower complexity requirements than the above-mentioned alternatives.

6.3 Experimental results

A number of experiments have been carried out with both synthetically generated optical flows and optical flows extracted from natural images. Figure 4 a) shows a synthetic affine optical flow containing five motion components having the same number of flow vectors. The parameter values have been randomly generated subject to some constraints in order to yield realistic flow fields. Components having the same sizes have been chosen to test the method in the critical conditions in which there is no dominant component. The correct segmentation shown in Figure 4 b) has been obtained in less than 300 cycles at each step through the image points. Note that no robust regression algorithm can deal with such data unless a preliminary and, in most cases, very problematic clustering stage is performed before the regressions. Figure 5 gives typical examples of the clusters obtained with the classical *k-medoid* algorithm [5] which is based on the minimization of the error distances between the vectors. Note that, although the correct number of components –which is usually unknown– has been used, the resulting clusters differ considerably from the actual motion components. The three reported segmentations correspond to different threshold values that have been used to define the set of candidate centroids based on the histogram of the vector components [5]. Thus, even if the correct number of components is known and the thresholds are tuned to have reasonable computational requirements, the clustering stage can yield a misleading decomposition.

In all tests with randomly chosen motion parameters without noise, the MIN PCS-based algorithm determined the correct segmentation as well as the actual motion parameters in less than 300 cycles at each step. For motion components of different sizes, an even smaller number of cycles of the thermal AMS procedure is needed. Figure 6 a) shows the same flow structure with the addition of pseudo-random Gaussian noise with a variance of about 20% of the average data dynamic which corresponds to a SNR of 15dB. As shown in Figure 6 b), the relevant structure of the scene is extracted but some small consistent subsystems appear due to the noisy flow vectors. The motion estimate for the large systems is nevertheless rather accurate and the spurious subsystems can be easily detected due to their (small) size and/or to their poor spatial coherence. In fact, the corresponding noisy flow vectors could even be corrected by performing an additional optical flow estimation based on the given segmentation.

Some experiments have also been carried out to compare these results with those provided by the standard HT. Correct segmentations are obtained for fine quantization of the parameter space, that is, at the cost of much higher computational complexity and storage requirements. Coarser quantizations yield partially incorrect segmentations with inaccurate estimates of the motion parameters. A more detailed experimental comparison is omitted for lack of space and will be reported elsewhere.

To assess the performance of our approach for fields that are affected by severe noise and errors, we have also experimented with optical flows extracted from natural images. It is worth emphasizing that no post-processing has been performed to refine the results using spatial coherence information or the grey level pixel values. Since noisy vectors yield small spurious consistent subsystems with no relevant modeling information, the consistent subsystems whose size is smaller than a certain threshold are discarded and are assigned to the relevant consistent subsystem whose solution gives the smallest violation. Selecting a value for this threshold, turns out not to be a critical issue. Large values may prevent the detection of small size moving objects, while smaller ones may yield small motion components with a poor spatial coherence. Note that the number of relevant consistent subsystems in the resulting partition provides an estimate of the number of motion components.
The results reported in Figures 7 and 8 for two test video sequences have been obtained by performing only 50 cycles of the thermal AMS procedure at each step. According to our experiments, the larger the number of data points in each submodel the smaller the number of cycles that are required in order to extract the corresponding close-to-maximum consistent subsystems. The *Mobile & Calendar* sequence (Figure 7) has been partitioned into three components which are correctly segmented with the exception of some errors in low-gradient areas that, for instance, prevent the segmentation of the rolling ball. The *Flower Garden* sequence (Figure 8) proves to be harder to segment due to the large optical flow estimation errors and due to the motion of the camera combined with the panning. Nevertheless, our MIN PCS-based algorithm succeeds in partitioning the image into the three main motion components that correspond to a first plane (the tree), a medium plane and a background. In both cases, a threshold value of 5% of the data point has been used to discard spurious consistent subsystems (larger values led to the same number of relevant motion components).

All experiments we have carried out indicate that our approach yields reliable *high level* segmentations at a very low computational cost. More accurate segmentations can then be obtained by using model-based approaches and a layered motion representation scheme such as in [40, 41]. Indeed, the spatial segmentation and the corresponding motion parameter values obtained with our method provide a very good starting point for a layered representation. By avoiding the necessary initial guesses or trial and error procedures, this can reduce very significantly the severe computational load which is the main drawback of this class of methods.

Finally, note that spatial constraints based on static intensity cues similar to the ones discussed in [36], can be easily incorporated into the MIN PCS-based framework. If there is strong evidence that nearby flow vectors must be assigned to the same submodel, we can just consider the corresponding blocks of 4 inequalities (7) as larger blocks. Situations where the evidence is weaker could then be taken into account by assigning a certain probability to any pair of neighboring flow vectors and, whenever one of the flow vectors is selected for a thermal AMS update, the update would also be performed with respect to its neighboring groups with the corresponding probability.

7 Concluding remarks

We have proposed a general combinatorial optimization approach to extract piecewise linear structure from nonlinear data. The problem of piecewise linear model estimation is formulated as that of partitioning inconsistent linear systems into a minimum number of consistent subsystems. Given the worst-case complexity of this MIN PCS problem and the type of problems we are interested in, we have presented a simple greedy algorithm which provides good approximate solutions in a short amount of time. This method enables simultaneous determination of the piecewise decomposition and of the parameter values of each linear submodel without requiring any a priori assumption on the discontinuities. By varying a single error tolerance parameter, one can guarantee different levels of accuracy.

The potential of our approach has been demonstrated on the challenging problem of 2-D multiple motion analysis from pre-computed optical flow fields. The experimental results indicate that it overcomes the serious limitations of some classical alternatives and compares favorably in terms of complexity and performance. In particular, it provides at a very low computational cost the number of motion components and the spatial location of the motion components even in the presence of severe noise and flow estimation errors.

The new approach can be applied to a variety of other problems for which piecewise linear models are appropriate. In [47] it is, for instance, adapted to the problem of piecewise linear modeling of time series. An application to line detection is underway. Since higher degree polynomials can be viewed as linear functions with respect to their coefficients, the approach can also be extended to the estimation of piecewise polynomial models with submodel of bounded degree.

Acknowledgments

The authors would like to thank Leslie Trotter Jr. for helpful discussions and Ramin Zabih for pointing out some references related to this work.

Appendix

Proof of theorem 1

To prove that MIN PCS is NP-hard we show that it is NP-complete to decide whether inconsistent systems can be partitioned into two consistent subsystems.

This decision problem clearly belongs to the class NP since, given any instance $(A, \mathbf{b})$ of MIN PCS and any bipartition of the linear system, one can check in polynomial time if both subsystems are consistent using any appropriate method to solve consistent linear systems.

NP-completeness is established by giving a polynomial time reduction from the classical NP-complete PARTITION problem that is defined as follows [13]. Given a finite set $S = \{1, \ldots, n\}$ and a size s_j for each $j \in S$, is there a subset $S' \subseteq S$ such that $\sum_{j \in S'} s_j = \sum_{j \in S - S'} s_j$?

Let $(S, \{s_1, \ldots, s_n\})$ be an arbitrary instance of the PARTITION problem. We construct a particular instance $(A, \mathbf{b})$ of MIN PCS such that the answer to the former one is affirmative (i.e. there exists a subset such that $\sum_{j \in S'} s_j = \sum_{j \in S - S'} s_j$) if and only if the answer to the latter one is also affirmative (i.e. it is possible to find a partition into two consistent subsystems).

The idea is to construct an inconsistent system containing one variable x_j for each $j \in S$. The system is composed of the equation

$$\sum_{j=1}^{n} s_j x_j = 0 \tag{8}$$

as well as of the $2n$ equations

$$x_j = 1 \quad \text{for} \quad j = 1, \ldots, n, \tag{9}$$

$$x_j = -1 \quad \text{for} \quad j = 1, \ldots, n. \tag{10}$$

Since for each j, $1 \leq j \leq n$, the equations of types (9) and (10) are contradictory, the overall system with $2n + 1$ is inconsistent. Clearly, such a system can always be partitioned in at most three consistent subsystems: one for equation (8), one for the equations of type (9) and one for those of type (10).

If the considered instance of PARTITION has a positive answer and $(S', S - S')$ is any bipartition of S such that $\sum_{j \in S'} s_j = \sum_{j \in S - S'} s_j$, the vector $\mathbf{x}$ defined by

$$x_j = \begin{cases} 1 & \text{if } j \in S' \\ -1 & \text{otherwise} \end{cases}$$

satisfies equation (8) and exactly n equations of types (9) and (10). The other n equations of types (9) and (10) are obviously consistent. The overall system can thus be partitioned into two consistent subsystems, one containing $n + 1$ equations and the other one n equations.

Conversely, suppose that the system (8)–(10) can be partitioned into two consistent subsystems and that $\mathbf{x}^1$ and $\mathbf{x}^2$ are two corresponding solutions. By construction, each of them satisfies exactly n equations of types (9)–(10) and hence has -1 or 1 components. If $\mathbf{x}^1$ denotes the solution which satisfies also equation (8), the bipartition of S defined by $j \in S'$ if and only if $x_j^1 = 1$ satisfies the relation $\sum_{j \in S'} s_j = \sum_{j \in S - S'} s_j$.

Sub-optimality of the greedy strategy

The following small example shows that even if maximum consistent subsystems are available at each step, the greedy strategy is not guaranteed to yield minimum partitions.

Example 3 Consider the system:

$$\begin{aligned} x_1 + x_2 &= 0 && (11) \\ x_1 - x_2 &= 0 && (12) \\ x_2 &= 1 && (13) \\ x_2 &= 2 && (14) \end{aligned}$$

whose largest consistent subsystems include two equations. Clearly, there are five corresponding solutions: $(0, 0)$ satisfies equations (11-12), $(1, 1)$ equations (12-13), $(2, 2)$ equations (12) and (14), $(-1, 1)$ equations (11) and (13) and $(-2, 2)$ equations (11) and (14). If one starts with the maximum consistent subsystem composed of equations (11-12), then two additional subsystems are needed in the partition because the two remaining equations are inconsistent. Now a minimum partition includes only two consistent subsystems, for instance equations (12) and (14) on one side and equations (11) and (13) on the other side.

Annealing schedule

As discussed in [20, 16, 18] for the thermal perceptron procedure, the choice of the initial temperature T_0 and the annealing schedule can have a significant impact on the quality of the solutions. Ideally, T_0 should be of the same order of magnitude as the typical values of the violations v^k. Clearly, the average $<\mid v \mid>= \sum_{k=1}^{p} \mid v^k \mid /p$ provides a reasonable estimate of the typical value of the v^k for a given $\mathbf{x}$. Since the $\mathbf{x}_i$ and therefore the v^k can vary considerably during the relaxation procedure, T_0 must be updated accordingly. To try to keep T_0 close to $<\mid v \mid>$ and to attenuate oscillations, one starts with $T_0 = \alpha <\mid v \mid>$, where the average is evaluated for the initial guess $\mathbf{x}_0$, and at the beginning of each random cycle one sets $T_0 = \beta(T_0 + <\mid v \mid>)$, where the average is for the current $\mathbf{x}_i$. Updating T_0 at the beginning of each random cycle corresponds to rescaling t with respect to the current $<\mid v \mid>$ without affecting the decrease of t given by γ. In practice, $\mathbf{x}_0$ is randomly generated, typical values of α and β are respectively 2 and 3/2, and γ is linearly decreased from T_0 to 0 over the C cycles ($\gamma = 1 - \frac{c}{C}$). See [10] for more details.

References

[1] M. Bertero, T. Poggio, and V. Torre. Ill-posed problems in early vision. *Proceedings of the IEEE*, 76(8):869–889, August 1988.

[2] J. Marroquin. Regularization theory and low-level vision. In M. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, pages 800–804, Cambridge, MA, 1995. MIT Press.

[3] T. Poggio, V. Torre, and C. Koch. Computational vision and regularization theory. *Nature*, 317:314–319, September 1985.

[4] D. Terzopoulos. Regularization of inverse visual problems involving discontinuities. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-8(4):413–424, July 1986.

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

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

[7] J. Illingworth and J. Kittler. A survey of the Hough Transform. *Comput. Vison Graphics Image Process.*, 44:87–116, 1988.

[8] V. F. Leavers. Which Hough Transform? *CVGIP: Image Understading*, 58(2):250–264, 1993.

[9] G. Adiv. Determining three-dimensional motion and structure from optical flow generated by multiple moving objects. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-7(4):384–401, July 1985.

[10] M. Mattavelli. *Motion analysis and estimation: from ill-posed discrete inverse linear problems to MPEG-2 coding*. PhD thesis, Communication Systems Division, Swiss Federal Institute of Technology, Thesis No. 1597, Lausanne, 1997.

[11] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth. Occam's razor. *Information Processing Letters*, 24:377–380, 1987.

[12] P. Crescenzi and V. Kann. A compendium of np-optimization problems. *Available at: http:// www.nada.kth.se /ṽiggo /problemlist /compendium.html*, 1995.

[13] M. R. Garey and D. S. Johnson. *Computers and Intractability: a guide to the theory of NP-completeness*. W. H. Freeman and Company, San Francisco, 1979.

[14] C. H. Papadimitriou and K. Steiglitz. *Combinatorial optimization algorithms and complexity*. Prentice Hall, New York, 1982.

[15] J. Ryan. Transversals of IIS-hypergraphs. *Congressus Numerantium*, 81:17–22, 1991.

[16] 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.

[17] E. Amaldi and V. Kann. The complexity and approximability of finding maximum feasible subsystems of linear relations. *Theoretical Computer Science*, 147(1-2):181–210, 1995.

[18] E. Amaldi, C. Diderich, and R. Hauser. On the probabilistic and thermal perceptron training algorithms. Technical report, 1997. Manuscript available at http://www.cs.cornell.edu/ Info/People/amaldi/amaldi.html.

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

[20] M. Frean. A "thermal" perceptron learning rule. *Neural Computation*, 4(6):946–957, 1992.

[21] S. Agmon. The relaxation method for linear inequalities. *Canadian Journal of Mathematics*, 6:382–392, 1954.

[22] T. S. Motzkin and I. J. Schoenberg. The relaxation method for linear inequalities. *Canadian Journal of Mathematics*, 6:393–404, 1954.

[23] Y. Censor. Row-action methods for huge and sparse systems and their applications. *SIAM Reviews*, 23:444–466, 1981.

[24] J. Telgen. On relaxation methods for systems of linear inequalities. *European Journal of Operational Research*, 9:184–189, 1982.

[25] R. Greer. *Trees and Hills: Methodology for Maximizing Functions of Systems of Linear Relations*, volume 22 of *Annals of Discrete Mathematics.* Elsevier science publishers B.V., Amsterdam, 1984.

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

[27] H. Maitre. The adaptive Hough Transform. *Traitement du signal*, 2, 1985.

[28] J. Illingworth and J. Kittler. The adaptive Hough Transform. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, PAMI-9(5):690–698, September 1987.

[29] 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.

[30] L. Xu, E. Oja, and P. Kultanen. A new curve detection method: Randomized Hough Transform. *Pattern Recognition Letter*, 11 (5):334–344, 1990.

[31] 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.

[32] 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.

[33] T. M. Chin, , M. R. Luettgen, W. C. Karl, and A. S. Willsky. An estimation theoretic prespective on image processing and the calculation of optical flow. In I. Sezan and R. L. Lagendijk, editors, *Motion Analysis and Image Sequence Processing*, chapter 2. Kluwer Academic Publishers, Boston, USA, January 1993.

[34] E. Dubois and J. Konrad. Estimation of 2-D motion fields from image sequences with application to motion compensated processing. In I. Sezan and R. L. Lagendijk, editors, *Motion Analysis and Image Sequence Processing*, chapter 3. Kluwer Academic Publishers, Boston, USA, January 1993.

[35] 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.

[36] Y. Weiss. A unified mixture framework for motion segmentation: incorporating spatial coherence and estimating the number of models. In *Proc. of the IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pages 321–326, 1996.

[37] J. B. Burt, R. Hingorani, and R. J. Kolczynski. Mechanism for isolating component patterns in the sequential analysis of multiple motion. In *Proc. of the IEEE Workshop on visual motion*, pages 187–193, Princeton, NJ, October 1991.

[38] J. Bergen, 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.

[39] N. Irani, B. Rousso, and S. Peleg. detecting and tracking multiple moving objects using temporal integration. In G. Sandini editor, editor, *Proceeding of the second European conference on computer vision*, pages 282–287, S. Margherita Ligure, Italy, May 1992.

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

[41] 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.

[42] Y. Weiss and E. H. Adelson. Motion estimation and segmentation using a recurrent mixture of experts architecture. In F. Girosi et al., editor, *Neural Network for Signal Processing V, Proceedings of the 1995 IEEE Workshop, Cambridge,Massachussetts*, pages 293–302, New York, NY, August 31, September 2 1995. IEEE Signal Processing Society.

[43] S. Ayer. *Sequential and competitive methods for the estimation of multiple motion*. PhD thesis, Department of Electrical Engineering, Swiss Federal Institute of Technology, Thesis No. 1405, 1996.

[44] Y. Weiss. Smoothness in layers: motion segmentation using nonparametric mixture estimation. In *Proc. of the IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pages 520–527, 1997.

[45] P. Bouthemy and E. Francois. Motion segmentation and qualitative dynamic scene analysis from an image sequence. *International journal of computer vision*, 10(2):157–182, 1993.

[46] M. Mattavelli and E. Amaldi. Using perceptron-like algorithms for the analysis and parameterization of object motion. In F. Girosi et al., editor, *Neural Network for Signal Processing V, Proceedings of the 1995 IEEE Workshop*, pages 303–312, Cambridge,Massachussetts, Aug. 31, Sept. 2 1995. IEEE Signal Processing Society.

[47] M. Mattavelli, J.M. Vesin, E. Amaldi, and R. Grueter. A new approach for piecewise linear modeling of time series. In *Proceedings of the 1996 Seventh IEEE Digital Signal Processing Workshop*, pages 502–505, Loen, Norway, September 1-4 1996.

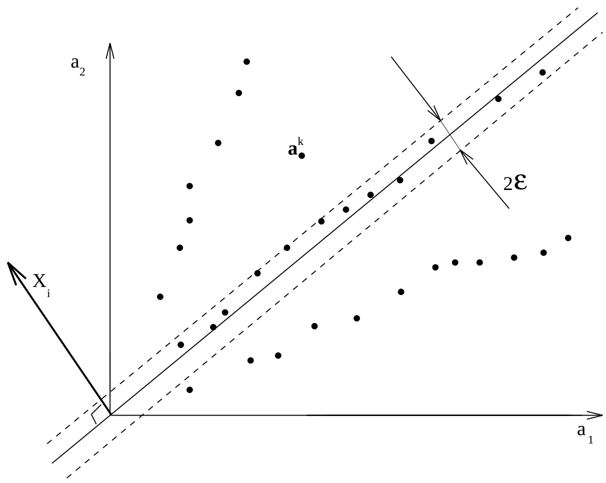


Figure 1: Geometric interpretation in coefficient space of the MIN PCS problem with pairs of complementary inequalities where $b_k = 0$, $1 \leq k \leq p$. For arbitrary b_k the hyperslabs do not necessarily contain the origin.

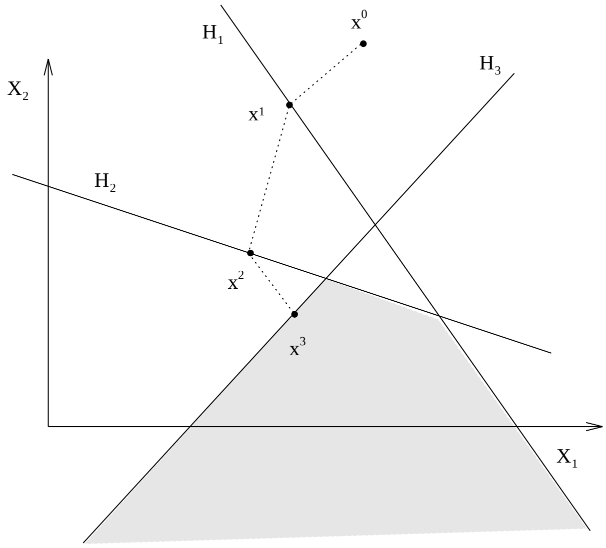


Figure 2: Geometric interpretation in parameter space of the AMS relaxation method for inequalities with $\lambda_i = 1$.

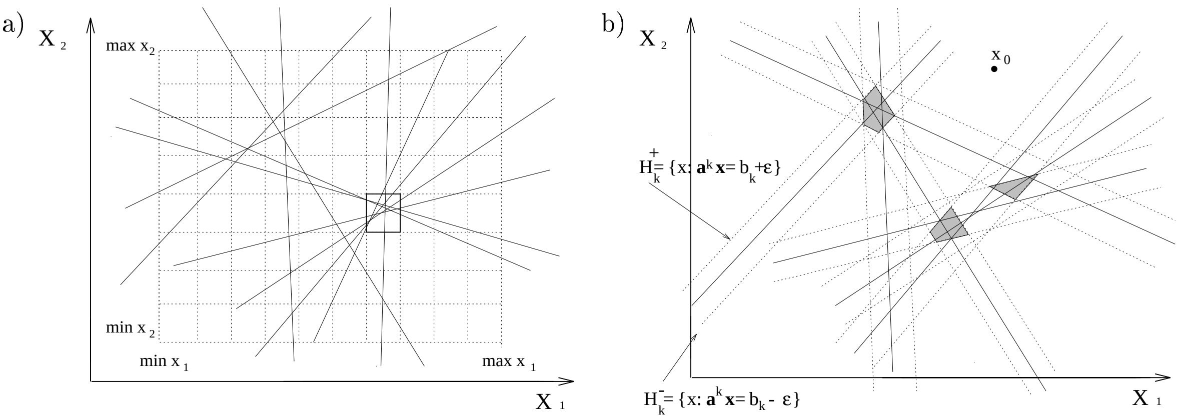


Figure 3: 2-D representation of the HT and MIN PCS-based strategies in parameter space. a) In the standard HT applied to linear models, each data point defines a hyperplane that is marked by incrementing all hypercells that intersect with it. Hypercells are then scanned to find a peak of the counter values, i.e., a hypercell consistent with a maximum number of hyperplanes. b) In our method, for each data point the hyperslab defined by the pair of corresponding complementary inequalities 2 is considered. Starting from an arbitrary $\mathbf{x}_0$, the current estimate is updated according to the thermal AMS procedure (similar to Figure 2) so as to look for an $\mathbf{x}$ contained in the largest number of hyperslabs. For clarity, those regions are shown in gray.

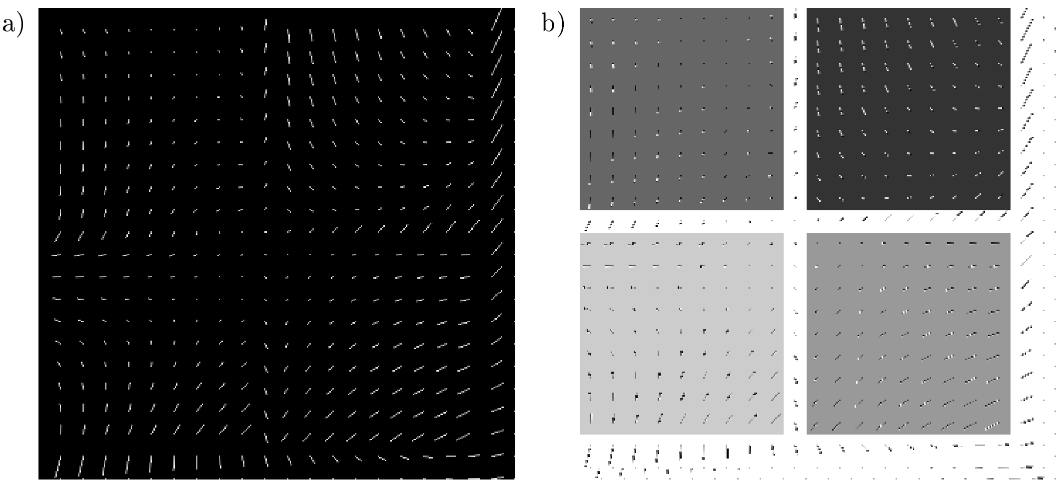


Figure 4: a) Optical flow field corresponding to five synthetically generated affine transformations. b) Segmentation of the field into the affine motion components, each gray level corresponds to a consistent subsystems.

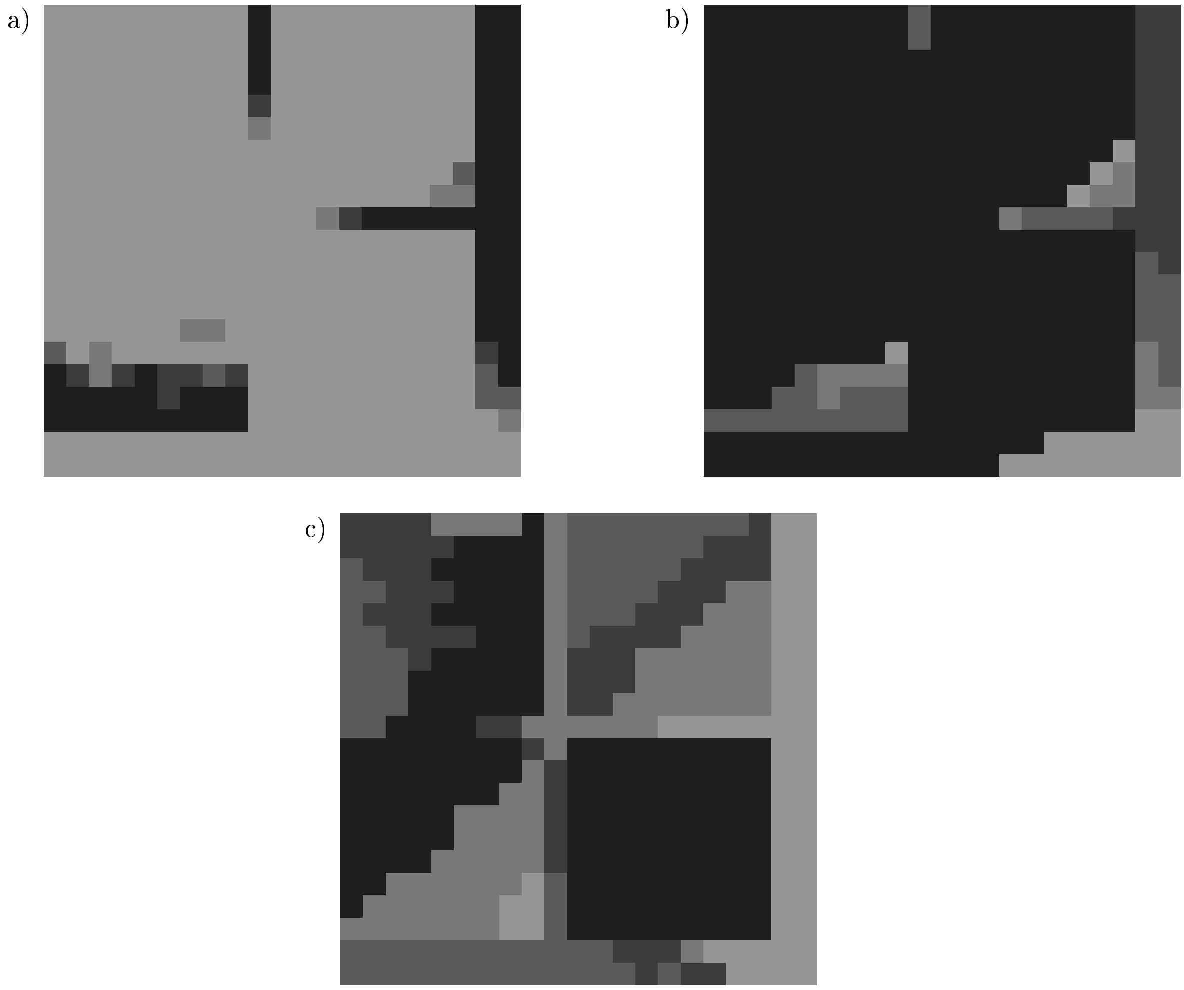


Figure 5: k-medoid clustering algorithm applied to the optical flow field of Figure 4 a). The segmentations in a), b) and c) are obtained for different values of the threshold used to select the histogram peaks indicating the centroid candidates. Although the correct number of components is given, the results are very poor and never approximate the sought partition.

a) b)

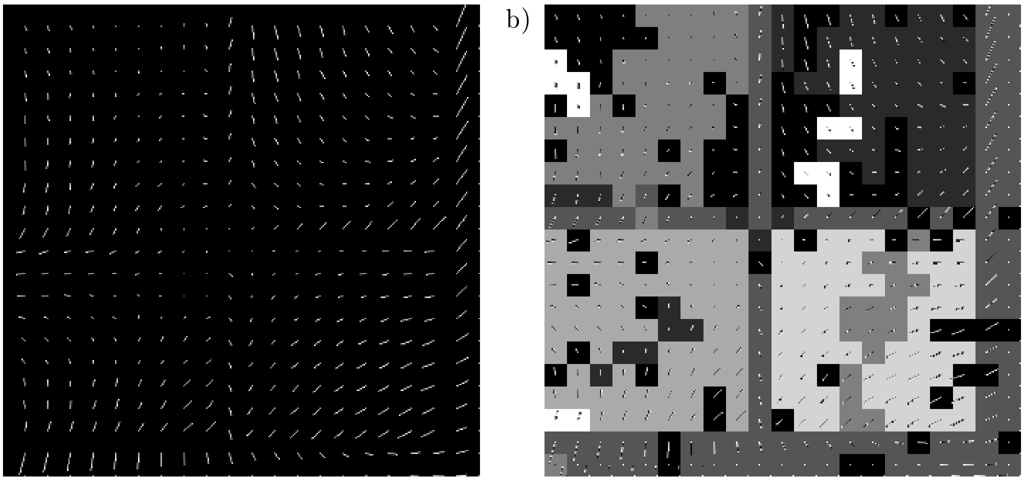
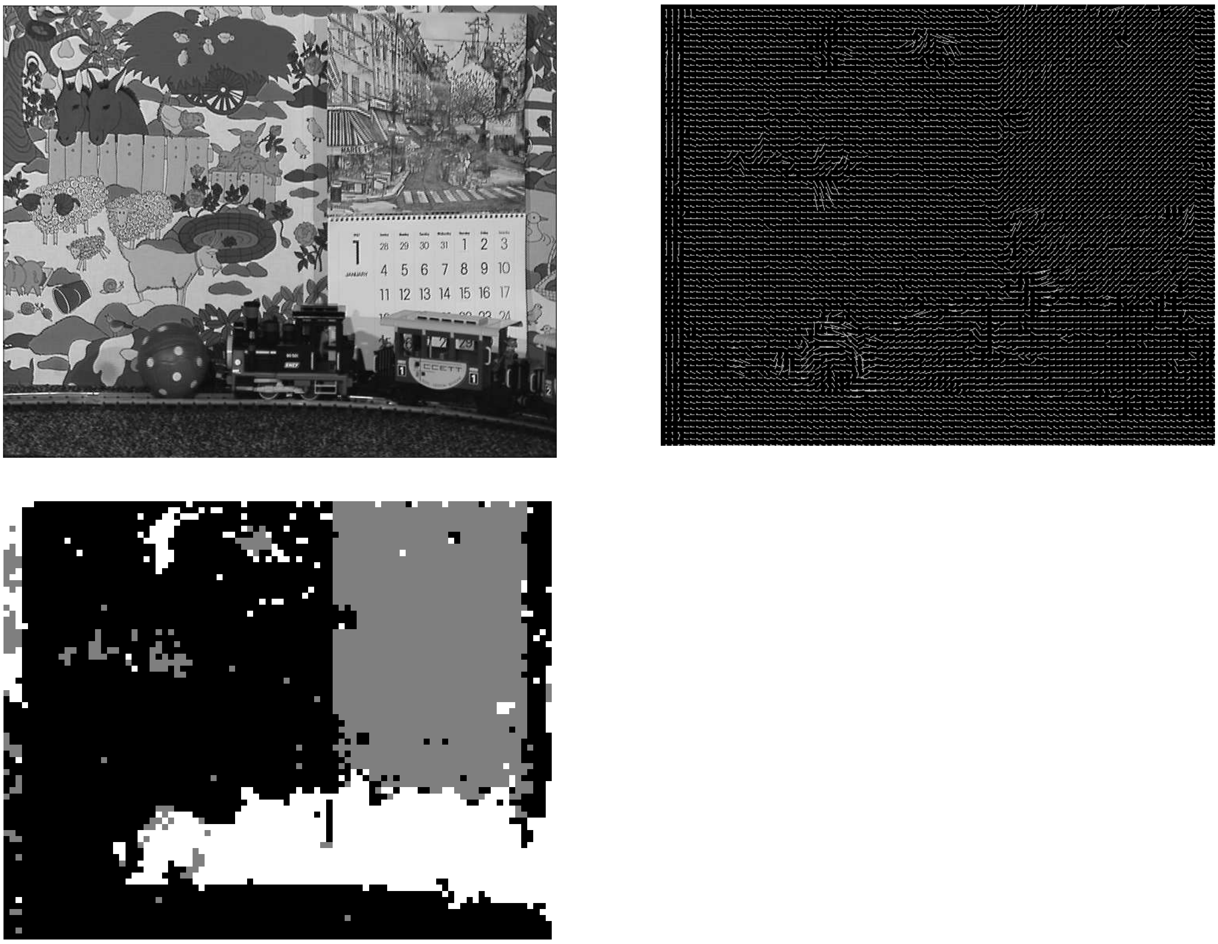


Figure 6: a) Optical flow field composed of five synthetically generated affine transformations and superimposed additive Gaussian noise (SNR = 15 dB). b) Segmentation of the noisy field into affine motion components, each gray level corresponds to a consistent subsystem.

a) b)



c)

Figure 7: *Mobile & Calendar* sequence. The apparent displacements of the objects are due to: a left to right panning of the camera, an upward displacement of the calendar and a right to left displacement of the train and the ball. a) First image of the sequence. b) Optical flow field obtained with a multigrid correlation technique [31]. c) Segmentation of the field into motion components, each gray level corresponds to a consistent subsystem.

a) b) c)

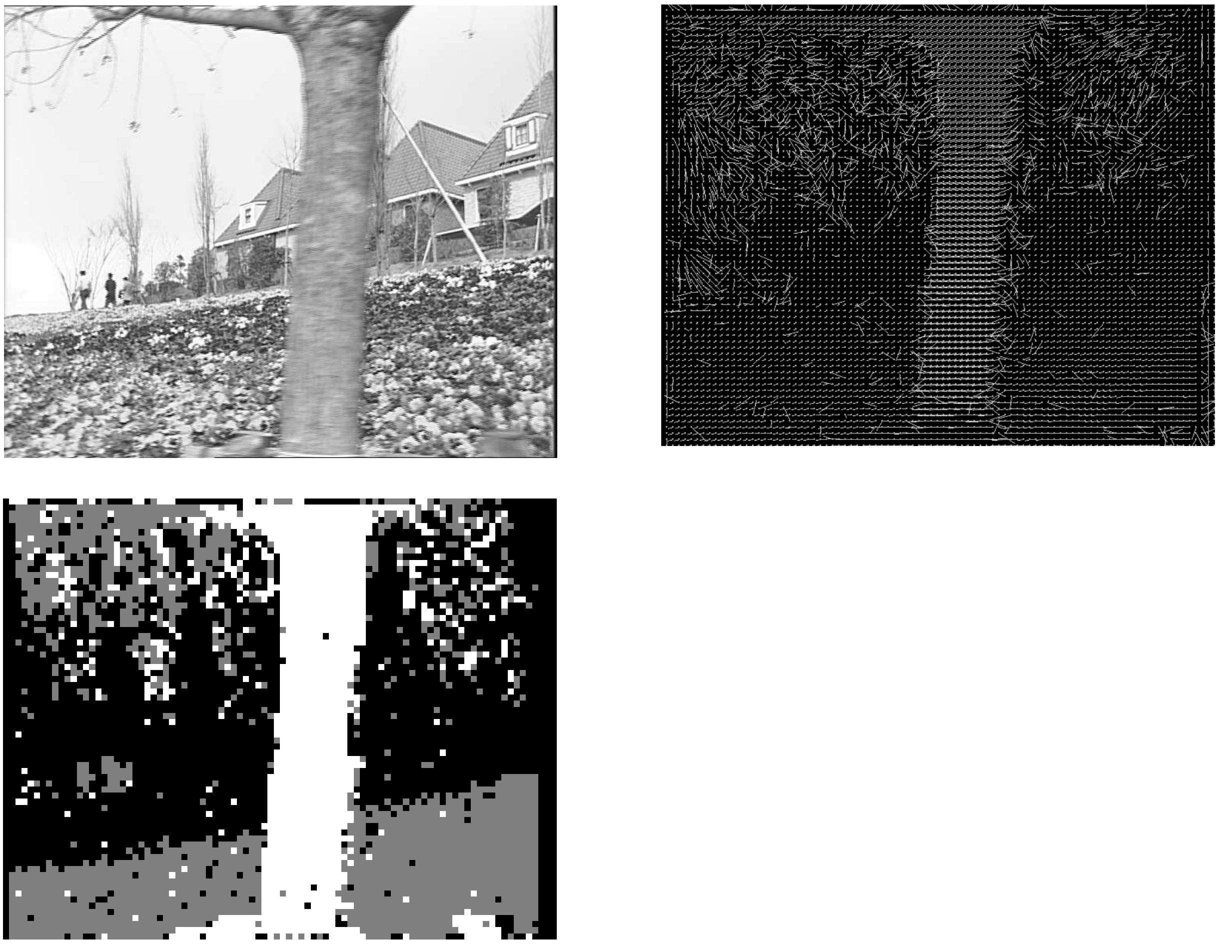


Figure 8: *Flower Garden* sequence. The apparent displacements of the scene elements are due to a left to right translation of the camera which induces a 3-D displacement. The apparent motion of the different elements of the scene depends on the distance from the camera objective. a) First image of the sequence. b) Optical flow field obtained with a multigrid correlation technique [31]. c) Segmentation of the field into motion components, each gray level corresponds to a consistent subsystem.