Disparity-Based Segmentation of Stereoscopic Foreground/Background Image Sequences

Til Aach and André Kaup

Abstract— In this contribution, we describe a method for displacement estimation in stereoscopic images, which is closely coupled with a segmentation of the pictures into homogeneously displaced regions. The technique is driven by a statistical optimization criterion which assesses the quality of the disparity estimate *and* of the segmentation, thus improving both of these simultaneously. In addition, the optimization criterion explicitly takes occluded areas into consideration. With the additional help of two constraints, this enables the algorithm to locate regions corresponding to occlusions accurately.

Keywords— Region based image coding, image analysis, disparity estimation, segmentation, Gibbs/Markov random fields, deterministic relaxation.

I. INTRODUCTION

Region oriented image coding strategies generally require that the images be partitioned into regions which should correspond to the objects in the depicted scene, or at least to in some respect homogeneous parts of these objects. The types of clues used to carry out this segmentation depend on the type of image material to be processed. In the case of moving video, segmentation is usually based on three-dimensional object motion in the observed scene or on apparent two-dimensional motion in the image plane. When objects moving in front of stationary background have to be delineated, it may suffice to just detect motion without a quantitative analysis [1]. Otherwise, the generally unknown motion has to be estimated prior to or during the segmentation process [2], [3], [4]. Applications of such segmentation techniques are, of course, not limited to data reduction, they are also of interest for tasks in image analysis (e.g. [5]).

In stereoscopic sequences, the disparity between the two stereo channels provides additional clues for segmentation. The reliability of motion-based segmentation may thus be improved by e.g. subjecting motion estimation to stereoscopic constraints [6], or by joint estimation of disparity and motion [7]. Additionally, taking into account disparity means that the segmentation results become less dependent on the sometimes insufficient amount of object motion [8], so that problems like falsely loading intensities of temporarily slowly moving objects into background memory (cf. [9]) can be avoided.

Dense disparity fields for coding applications can e.g. be estimated by matching image intensities [10], [8] or by differential approaches [11]. As an inverse problem, however, disparity estimation is ill posed [12], [13]. Thus, external smoothness constraints are necessary to restrict the space of possible solutions. These smoothness constraints may apply to the disparity fields ('surfaces') as well as to the region boundaries of a segmentation associated with a disparity field. Care should be taken that surface smoothness is not imposed on a disparity field as a whole, but only to regions belonging to individual objects, so as not to unduly smooth across region boundaries (cf. 'weak' constraints [14]). Finally, identification and proper treatment of occlusions must be a vital part of the estimation and segmentation procedure. Occluded areas depict parts of the scene which are visible for one camera only, what means that they have no proper counterpart in the other camera's image [1]. Since the attempt to attribute a genuine disparity to these areas is wrong as such, the corresponding locations must be uniquely marked as occluded. 'Structural' rules like ordering constraint [15] or uniqueness [16] may be exploited for detection of occlusions.

In this contribution, we describe a technique which combines disparity estimation with a segmentation of stereo images into regions of constant disparity. Our method is hence especially well suited for stereo image pairs of scenes in which each object is characterized by an individual displacement. This can usually be assumed if the dimensions-in-depth of the visible object parts are small compared to their distance ('depth') from the camera set. Specifically, we concentrate here on scenes where a speaker is placed in front of background. For coding applications, a segmentation into foreground and background enables one to treat these regions differently, e.g. by allocating more bits for the person at the expense of the subjectively less interesting background.

The surface smoothness constraint appears here in its most stringent form since it requires displacements inside each region to be identical, thus uniquely relating a partition to any disparity field. Smoothness of the region boundaries is realized by modeling the partition as a Gibbs/Markov random field. The combined problem of disparity estimation and segmentation is tackled through the maximum a posteriori approach, i.e. we try to find the disparity field which is most probable to occur given the ob-

served image data. Detection and treatment of occlusions are directly integrated into this approach.

In the following we first describe the model underlying the stereo images and the properties we require the disparity fields to exhibit. Then, a criterion assessing the quality of disparity fields and their associated partitions is derived. This criterion also explicitly evaluates areas classified as occluded. Section III B explains a deterministic optimization procedure, and section III C describes how initial estimates for this iterative procedure can be obtained. We conclude by giving some results and a discussion.

II. THE STEREO IMAGE MODEL

Let us denote the grey levels of the observed picture pair by $Y_i = \{y_i(k)\}$, $k = 1, \ldots, N$, with N being the number of pixels in each image and $i = 1, 2$ indicating the two perspective views. We assume the grey level difference field $D_0 = \{d_0(k)\}$, $d_0(k) = y_1(k) - y_2(k)$ between the two views as being solely due to the underlying disparity field and to noise caused by the two cameras.

'True' disparity fields associated with the type of scenes we are dealing with here exhibit certain properties due to their physical world origin. In order to make our disparity estimates comply with these, the space of possible solutions is subject to the following constraints ([17], cf. also [18], [19]):

> *Uniqueness:* For each pixel in one image Y_1, there either exists a one-to-one correspondence with one point in the other image Y_2 and vice versa, or the pixel is assigned to uncovered background/occlusions.
>
> *Piecewise Homogeneity:* The disparity inside regions is constant.
>
> *Boundary Smoothness:* Smooth region boundaries are to be preferred.

Uniqueness and piecewise homogeneity are used in a similar form in [19], however, they are formulated more strictly here. The uniqueness constraint as stated here requires an unambiguous two-way correspondence between two pixels linked by a disparity vector. To be more precise, let us assume a displacement vector $v(k)$ which points from, say, pixel $y_1(k)$ in image Y_1 to pixel $y_2(l)$ in the other image Y_2 (Fig. 1). Uniqueness requires a reverse vector to exist originating from $y_2(l)$ and pointing to $y_1(k)$. Since at most one disparity vector is admitted per pixel, the implications are twofold (see Fig. 1): firstly, the disparity field pointing from Y_1 to Y_2 uniquely determines the reverse field from Y_2 to Y_1. Secondly, it is straightforward to see that a disparity field which is not completely homogeneous must contain pixels to which no disparity complying with uniqueness can be assigned. These pixels have no counterparts in the other image, usually due to occlusions. In accordance with what was said above, they are marked by a pseudo disparity called 'NIL'. Uniqueness thus serves as a structural rule for the unambiguous detection of occluded areas.

The second constraint, piecewise homogeneity, relates a disparity field V to a segmentation Q by combining identically displaced pixels to regions. The third, rather 'soft' statement concerning region boundary smoothness will be
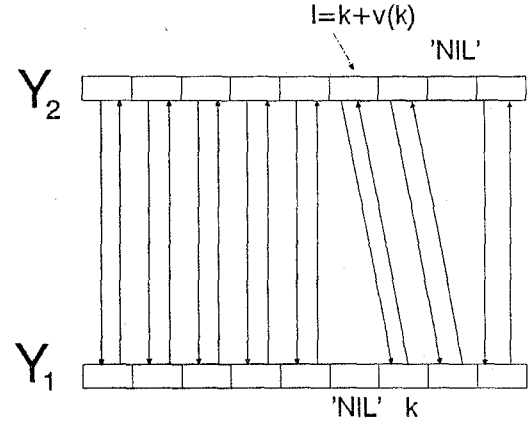

Fig. 1. The *uniqueness* rule illustrated for a 1D-case.

incorporated into the optimization criterion by an appropriate a priori probability density (pdf) reflecting a preference for smooth partitions.

III. MAP-APPROACH TO DISPARITY ESTIMATION

The disparity field $V = \{v(k)\}$ shall be estimated such as to maximize the a posteriori probability density $p(V|Y_1, Y_2)$. This can be restated as the maximization of the product $p(Y_1, Y_2|V) \cdot p(V)$, where $p(Y_1, Y_2|V)$ denotes the likelihood function of disparity fields V based on the observed image data, and $p(V)$ the a priori pdf for the disparity fields. The advantage of the MAP-approach is that through a proper specification of the a priori pdf $p(V)$ we can bias the estimation procedure towards solutions consistent with our prior expections. Because of its flexibility, we chose a Gibbs/Markov random field to model the disparity fields (cf. [20], [21], [22]). The a priori pdf $p(V)$ is thus given by

$$ p(V) = \frac{1}{Z_V} \cdot \exp\{-E(V)\} \ , \tag{1} $$

where Z_V is a normalization constant. $E(V)$ is a so-called *energy* term depending on V. The notion "energy" refers to statistical mechanics where this model, which favours low-energy states, is taken from. Its versatility is due to the fact that within certain limitations outlined in the Hammersley-Clifford theorem [22, p. 198], the energy may be specified almost arbitrarily to fit the particular problem encountered. Here, the aim is to find an energy term related to the boundary smoothness of the segmentation Q associated with V.

The boundary smoothness can be assessed by the number of border pixel pairs present in a partition Q. A border pixel pair – also termed *inhomogeneous clique* – consists of two horizontally, vertically or diagonally directly adjacent image points, which belong to different regions. In other words, border pixel pairs are situated across region boundaries. The number of border pixel pairs is hence lower when the boundaries in the partition are smoother (cf. e.g. [20], [21]). Distinguishing between horizontally/vertically and diagonally oriented border pixel pairs, we can specify the

energy $E(V)$ of the displacement estimate V giving rise to Q by

$$E(V) = n_B \cdot B + n_C \cdot C \ , \qquad (2)$$

where n_B denotes the number of horizontal and vertical border pixel pairs, and n_C that of the diagonal ones. The so-called *potentials* B and C are both positive, incurring the cost B or C on the solution for each occuring border pixel pair. Combining (2) with (1) results in an a priori pdf which assigns a higher probability of occurence to disparity fields which lead to smooth partitions than to other ones which do not.

With $p(V)$ specified, we now focus on the likelihood function $p(Y_1, Y_2 | V)$. The dependence of the likelihood function on Y_1 and Y_2 is modeled in terms of the displaced frame difference (DFD) $D(V) = \{d(k, v(k))\}$. Assuming that the influence of the underlying disparity field on the difference image D_0 is compensated by an estimate V, what remains is the influence of the camera noise which is modeled as being white and Gaussian distributed. The likelihood function is then given by [2]

$$p(Y_1, Y_2 | V) = \frac{1}{Z} \prod_{k=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left\{-\frac{d^2(k, v(k))}{2\sigma^2}\right\} \ , \quad (3)$$

with Z such that $\sum_{Y_1, Y_2} p(Y_1, Y_2 | V) = 1$. Replacing the unknown parameter σ^2 in (3) by its ML-estimator $\hat{\sigma}^2(V) = 1/N \cdot \sum_{k=1}^{N} d^2(k, v(k))$ yields

$$p(Y_1, Y_2 | V) = \frac{1}{Z} \left(\frac{1}{\sqrt{2\pi\hat{\sigma}^2(V)}}\right)^N \cdot \exp\left\{-\frac{N}{2}\right\} \ . \quad (4)$$

This expression can be maximized by determining V such that the mean square difference $\hat{\sigma}^2(V)$ is minimized, yielding a ML-estimate of the underlying disparity field. By combining (1) and (4), an expression proportional to the a posteriori density $p(V | Y_1, Y_2)$ can be formed. However, the specification of this density is not yet complete, since the above considerations do not provide any means for treating occlusions.

A. Treatment of Occlusions

As already described, pixels depicting occlusions can structurally be detected by invoking uniqueness. For these pixels no correspondence exists, so that instead of a genuine disparity they are formally marked by the pseudo disparity NIL. Pixels carrying the pseudo disparity NIL cannot, however, be assessed by (4), since no displaced pixel difference $d(k, v(k)) = NIL$ can be computed. To evaluate NIL-regions we decompose $p(Y_1, Y_2 | V)$ into $p_{NIL}(Y_1, Y_2 | V)$ assessing NIL-pixels and $p_G(Y_1, Y_2 | V)$, which covers all image points with a genuine displacement [17], [24].

[2] The likelihood function is thus being modelled proportional to $p(D(V) | V)$. Alternatively, one could follow the formulation of Konrad [23] by rewriting the maximization of $p(V | Y_1, Y_2)$ as the maximization of $p(Y_1 | V, Y_2) \cdot p(V | Y_2)$. If V is assumed as statistically independent from Y_2 [23, chapter 3], this leads to the same expression as our approach.

With N_{NIL} being the number of NIL-pixels, $p_G(Y_1, Y_2 | V)$ is given by

$$p_G(Y_1, Y_2 | V) \propto \left(\frac{1}{\sqrt{2\pi\hat{\sigma}^2(V)}}\right)^{N - N_{NIL}} \exp\left\{-\frac{N - N_{NIL}}{2}\right\} \quad (5)$$

with $\hat{\sigma}^2(V) = 1/(N - N_{NIL}) \sum_{v(k) \neq NIL} d^2(k, v(k))$. In order to integrate evaluation of NIL-pixels into the likelihood, we formally describe these by another random process which is also supposed to be Gaussian. We couple this formal process to the genuine one by assuming its formal variance as being proportional to the estimate $\hat{\sigma}^2$. We thus have

$$p_{NIL}(Y_1, Y_2 | V) \propto \left(\frac{1}{\sqrt{2\pi\hat{\sigma}_{NIL}^2}}\right)^{N_{NIL}} \exp\left\{-\frac{N_{NIL}}{2}\right\} \ ,$$

$$\hat{\sigma}_{NIL}^2 = F \cdot \hat{\sigma}^2(V) \ . \quad (6)$$

If the proportionality factor F is chosen greater than one, this formal difference process can be interpreted to have on the average a higher squared (displaced) grey level difference than the genuine one. This is reasonable since the pseudo disparity NIL should be placed only where all efforts to establish a correspondence fail.

Since the NIL-random process and the genuine one are mutually independent, $p(Y_1, Y_2 | V)$ is given by the product $p(Y_1, Y_2 | V) = p_G(Y_1, Y_2 | V) \cdot p_{NIL}(Y_1, Y_2 | V)$. Having thus integrated into the likelihood expression a term responsible for occlusions, the a posteriori pdf $p(V | Y_1, Y_2)$ finally is proportional to the product of (1), (5) and (6). Taking minus twice the logarithm of the resulting expression and dropping all additive constants, finding the MAP-estimate of V is equivalent to minimizing

$$K(V) = N \ln \hat{\sigma}^2(V) + N_{NIL} \ln(F) + 2 \cdot (n_B \cdot B + n_C \cdot C) \quad (7)$$

by variation of V. Minimization of $K(V)$ is a compromise between minimization of each one of its three components, each of which has its specific purpose: the first term, $N \ln(\hat{\sigma}^2)$, penalizes discrepancies between the input data and the found solution. The second term, $N_{NIL} \ln(F)$, incurs on the solution the cost $\ln(F)$ for each pixel classified as occluded, thus preventing undue spreading of NIL-regions. The third term, derived from the a priori pdf, is responsible for preferring displacement estimates which lead to smooth partitions.

B. Optimization by Deterministic Relaxation

Minimization of (7) is carried out by a deterministic relaxation of the ICM-type (cf. [25]). As an iterative algorithm, this displacement vector relaxation needs an initial estimate to start from. When a sequence of stereo image pairs is to be processed, it is reasonable to simply take the processing result for the previous frame pair as initialization for the next one, thus stepping through the sequence recursively. Gradual changes in time of the disparity for individual objects can be accommodated by performing a segment matching in a small search area centered around

the tip of the old disparities. Necessity for modifications then exists mainly at region borders (s. Fig. 4), hence, it is here where the displacement vector relaxation focuses.

The image field is scanned repeatedly, with a new scan direction for every scan. Whenever a pixel k situated at a region boundary is encountered, its displacement vector $v(k)$ is tentatively replaced by the displacement(s) of the region(s) adjacent to it. For a pixel next to a NIL-region, this also includes replacing its disparity by a tentative NIL. Each replacement operation simultaneously transfers pixel k to the corresponding neighbouring region, thus locally altering the region boundary's trajectory. For each replacement, (7) is recomputed. That vector (including the 'old' one) which minimizes (7) is retained. Recomputing (7) is considerably simplified by the local character of the replacement operations. However, since the new tentative displacement field(s) generated by replacing $v(k)$ by its neighbour(s) must comply with the uniqueness constraint, neighbouring displacements may be affected, too. These cases are listed in the appendix.

The relaxation converges as only operations which decrease (7) are performed. In practice, the relaxation may be terminated when the number of changed displacement vectors per scan has fallen to an insignificant value, e.g. 500 for 256 × 256 images. The fact that (7) is minimized only locally is no drawback when one relies upon such 'good' initial estimates as those formed by the previous processing results. Special measures, however, have to be taken if no such initialization exists, as it is e.g. the case for the first frame pair of a sequence.

C. Estimating Initial Displacement Fields

In cases where the above recursive approach is not applicable, e.g. after scene cuts in a sequence, or when single image pairs are to be processed, initial estimates have to be provided by separate strategies. A possibility near at hand for this purpose is to use block matching. However, since no global constraints like smoothness are brought to bear by this technique, the resulting displacement fields may exhibit large areas with wrong disparities which cannot be removed completely due to the deterministic nature of the above relaxation, even when restricting the block matcher's search space by the epipolar constraint. A strategy to avoid this pitfall could be to further reduce the number of disparity vectors to be tested by block matching using the phase correlation technique (cf. [8]). More appealing computationally is to implement a pyramid version of the displacement vector relaxation. Block matching is carried out on the lowest resolution level only, where the solution space is considerably decreased. After applying the relaxation, the estimation result is enlarged for the next higher resolution level. To account for the refined resolution, this must include a segment matching around the tip of the enlarged displacement vector of each region, before the relaxation can again be applied. This is repeated until the highest resolution level is reached.

Before the displacement vector relaxation starts, however, the initial estimate generated by the block matcher

has to be modified in order to comply with the uniqueness constraint. This means that, whenever two or more displacement vectors point towards the same location in the other image, only that one with lowest absolute displaced pixel difference is retained, the others are removed and replaced by the pseudo disparity NIL.

IV. RESULTS

In this section, we show processing results for 3 different videotelephony sequences consisting of frames sized 256×256 pels. The sequences were recorded with the stereo camera setup described in [8]: the cameras were mounted one vertically above the other, with their optical axes converging on the foreground (person) [3]. The partitions produced by the described algorithm consist of the regions person (foreground), background, and uncovered background ('NIL'). The region boundaries are in each case overlaid on the input images. Before processing started, the pictures of each pair were scaled to same global mean and variance.

First, Fig. 2 depicts the segmentation result for the first frame of sequence 1: left, we have the partition for the upper camera, with uncovered background above the person's head and shoulders. On the right hand side, the complement for the lower camera is depicted, where uncovered background occurs below the upper image edge, and below the person's ears. This segmentation was acquired by a pyramid consisting of the resolution levels 256 × 256 pels, 128 × 128 pels and 64 × 64 pels. On each level, the displacement vector relaxation was carried out in two phases, first with the rather low cost parameters $B = 0.5$, $C = 0.25$, $\ln(F) = 2$, and subsequently with the higher values $B = 5$, $C = 2.5$, and $\ln(F) = 5$. Hence, during the first phase, the image data term in (7) dominates resulting in preliminary estimates for V and Q which are particularly well adapted to the input data. However, since for this case the regularizing influence of the smoothness term in (7) is only slight, region borders may become very irregular. This effect is mended in the second phase, where the smoothing term is weighted higher. The above values for the parameters were determined experimentally, but they are not critical, and the described schedule was used in all our experiments.

The initialization was obtained by block matching on the lowest resolution level. To bring about the depicted results from this initialization, the displacement vector relaxation carried out 2856 vector replacements requiring 5.5 full scans[4]. Block matching and displacement vector relaxation took 88 seconds CPU-time on a VAXstation 3100 M76. The vertical component of the disparity field is given

---

[3] The test sequences used here were kindly provided by the author of [8], who for his investigations gave specific reasons for mounting the cameras vertically instead of horizontally. Note that the geometry of the camera setup nowhere entered the development of the algorithm described here, so that a horizontal setup can be used as well.

[4] We define a full scan to be equivalent to one raster scan on the finest resolution level of 256 × 256 pels, or to 4 raster scans on the next lower resolution level of 128 × 128 pels, and so on. The number of full scans as given above thus serves as a comparative measure of computational expense (cf. [26], [27]).

Fig. 2.   Region boundaries obtained by a multiresolution displacement vector relaxation, with initialization obtained by block matching on the lowest resolution level. Left: result for upper camera. Right: segmentation complement for the lower camera, computed from the upper one by uniqueness and piecewise homogeneity.
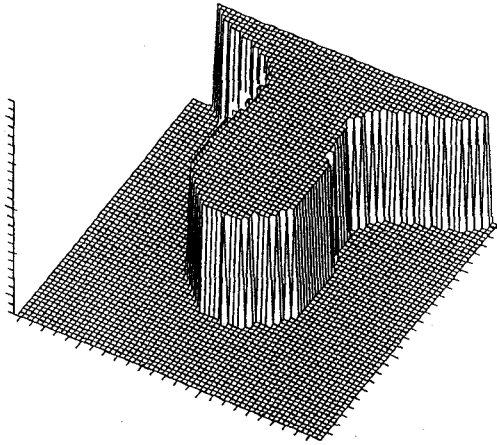

Fig. 3.   Vertical component of the disparity field associated with Fig. 2. To avoid 'holes' in the representation, the pseudo disparity 'NIL' has been replaced by the background's displacement.

in Fig. 3. The disparity of the foreground (person) was estimated to $(0,0)$ and that one of the background to $(1,12)$.

This example illustrates well how the smoothness term in (7) dominates in areas where it is difficult to separate different regions. This is e.g. the case along the person's throat, where the precise boundary even visually is hard to tell. Here, the third term of (7) ensures a rather smooth boundary instead of a ragged, 'uncertain' one.

The recursive adaptation of disparity estimation and segmentation along a sequence is illustrated in Fig. 4. The processing result for the 4th frame is adapted to the 6th one, requiring 34 seconds CPU-time.

Fig. 5 shows on its left hand side an initial segmentation obtained by block matching on the lowest resolution level of a pyramid. The right hand side gives the partition computed from this initial estimate by the relaxation, which carried out 4272 replacement operations during $5\frac{5}{8}$ full scans.

V. DISCUSSION

The described displacement vector relaxation combines estimation of stereo disparity fields with the segmentation of these fields into homogeneously displaced regions. One advantage of the proposed method is that these two tasks are not carried out independently of each other, since the optimization criterion (7) reflects the needs of both segmentation (using the Gibbs model) and displacement estimation (using the DFD). With respect to the accuracy of localizing the region boundaries, it is of further advantage that the algorithm works pixel by pixel (or rather, vector by vector), thus avoiding the use of a smearing measurement window. In addition to separating foreground from background, the algorithm also marks regions corresponding to occlusions accurately.

From a computational point of view it is appealing that only border pixels need to be considered. Furthermore, a possibility for a speedup is to implement a parallel version of the relaxation algorithm by 'synchronous updating' (cf. [25]).

The applications are so far limited to scenes containing objects which can be characterized by homogeneous disparity. Extensions require easing the constraints mentioned in section II towards allowing moderate variations of disparity inside regions, e.g. by modeling the disparity by parametric functions, or by more general models of mathematical physics, like membranes ([14], [11], [28]). For such more complicated scenes, however, finding discontinuities as well as a general treatment of occlusions are not yet solved (cf. [29]).

APPENDIX

In the following, we list 5 different situations into which the vector replacement operations in section III B may be classified, and illustrate how each one is treated (see Fig. 6).

   a) A genuine disparity $v(k)$ located at a boundary of one region is to be replaced by another genuine disparity of an adjacent region. Besides replacing $v(k)$, the target pixel (in image Y_2) of the replaced vector must be

Fig. 4. Recursive adaptation when processing a stereoscopic sequence, shown for the upper camera only. Left: boundaries of frame no. 4 overlaid on the new frame no. 6. Right: boundaries when the estimate shown left has been adapted to the actual frame by carrying out 2977 vector replacements in 7 full scans.
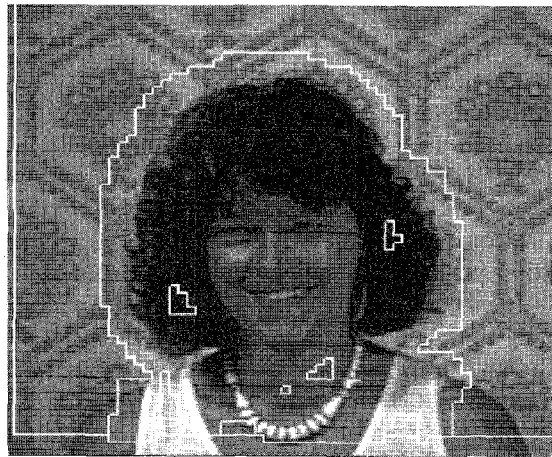

Fig. 5. Region boundaries (upper camera only) obtained by a multiresolution displacement vector relaxation, with initialization obtained by block matching. The segmentation computed by piecewise homogeneity from the initial estimate is shown left (enlarged from 64 × 64 pels resolution), and the refined one is depicted right. Measured displacements: $(-4, 12)$ for the background, and $(-5, 0)$ for the foreground.
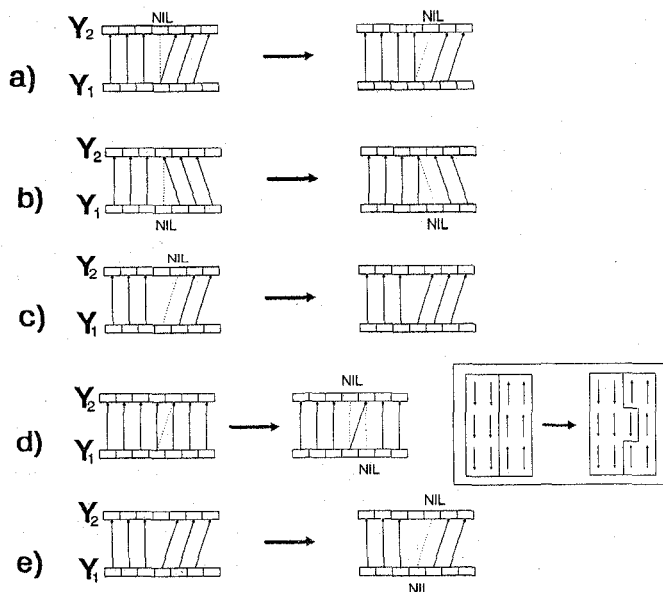

Fig. 6. Illustrations of the different situations which may arise during the vector replacement operations, for simplicity reduced to one dimension.

assigned the pseudo disparity NIL.

b) Pixel k carries the pseudo disparity NIL, which is to be replaced by the genuine disparity of an adjacent region. After the replacement operation, that vector which points to the same pixel in image Y_2 as the replacing one must be replaced by NIL, since otherwise the two vectors point to the same target pixel. When viewed in the reverse direction, i.e. from Y_2 to Y_1, this is identical to a).

c) Origin (in image Y_1) as well as the target point of the replacing vector carry the pseudo disparity NIL. In this case, no further pixels are affected.

d) Origin as well as target point for the replacing vector carry genuine disparities. Besides the replacement operation itself, this case combines the implications of a) and b). This situation arises when the disparities on either side of the region border are parallel and parallel to the border (see Fig. 6 d, right).

e) A genuine disparity is to be replaced by NIL. Here, both origin and target of the replaced genuine disparity are set to NIL.

REFERENCES

[1] C. Lettera and L. Masera, "Foreground/background segmentation in videotelephony", *Signal Processing: Image Communication*, vol. 1, pp. 181–189, 1989.

[2] M. Hötter and R. Thoma, "Image segmentation based on object oriented mapping parameter estimation", *Signal Processing*, vol. 15, pp. 315–334, 1988.

[3] H. Nicolas and C. Labit, "Global motion identification for image sequence analysis and coding", in *Proceedings ICASSP 91*, Toronto, Canada, May 1991, pp. 2825–2828.

[4] B. Choquet and D. Pelé, "Segmentation of apparent 2D motion fields", in *3rd International Conference on Image Processing and its Applications*, Warwick, UK, August 1988, IEE, pp. 280–284.

[5] E. Francois and P. Bouthemy, "Multiframe-based identification of mobile components of a scene with a moving camera", Tech. Rep. 564, IRISA, 1990.

[6] H. Nicolas, A. Tamtaoui, and C. Labit, "Estimation de descripteurs de mouvement: Application au codage de séquences monoculaires et stéréoscopiques", in *Proceedings 13ième Colloque GRETSI*, Juan-Les-Pins, France, September 1991, pp. 873–876.

[7] A. Tamtaoui and C. Labit, "Constrained disparity and motion estimators for 3DTV image sequence coding", *Signal Processing: Image Communication*, vol. 4, pp. 45–54, 1991.

[8] M. Waldowski, "A new segmentation algorithm for videophone applications based on stereo image pairs", *IEEE Transactions on Communications*, vol. 39, no. 12, pp. 1856–1868, 1991.

[9] S. C. Brofferio, "An object-background image model for predictive video coding", *IEEE Transactions on Communications*, vol. 37, no. 12, pp. 1391–1394, 1989.

[10] H. Harasaki, M. Yano, and T. Nishitani, "Background separation/filtering for videophone applications", in *Proceedings ICASSP 90*, Albuquerque, NM, USA, April 1990, pp. 1981–1984.

[11] R. March, "Computation of stereo disparities using regularization", *Pattern Recognition Letters*, vol. 8, pp. 181–187, 1988.

[12] M. A. Bertero, T. Poggio, and V. Torre, "Ill-posed problems in early vision", *Proceedings of the IEEE*, vol. 76, no. 8, pp. 869–889, 1988.

[13] T. Poggio, "Early vision: From computational structure to algorithms and parallel hardware", *Computer Vision, Graphics, and Image Processing*, vol. 31, pp. 139–155, 1985.

[14] A. Blake and A. Zisserman, *Visual Reconstruction*, The MIT Press, Cambridge, London, 1987.

[15] J. J. Little and W. E. Gillet, "Direct evidence for occlusion in stereo and motion", in *Lecture Notes in Computer Science 427*, O. Faugeras, Ed., Antibes, France, April 1990, ECCV 90, pp. 336–340, Springer Verlag.

[16] S. I. Olsen, "Stereo correspondence by surface reconstruction", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 3, pp. 309–315, 1990.

[17] T. Aach, A. Kaup, and R. Mester, "Combined displacement estimation and segmentation of stereo image pairs based on Gibbs random fields", in *Proceedings ICASSP 90*, Albuquerque, NM, USA, April 1990, pp. 2301–2304.

[18] D. Marr, *Vision*, Freeman, San Francisco, 1982.

[19] D. Marr and T. Poggio, "Cooperative computation of stereo disparity", *Science*, vol. 194, pp. 283–287, 1976.

[20] H. Derin and W. S. Cole, "Segmentation of textured images using Gibbs random fields", *Computer Vision, Graphics, and Image Processing*, vol. 35, pp. 72–98, 1986.

[21] R. Mester and U. Franke, "Statistical model based image segmentation using region growing, contour relaxation and classification", in *Proceedings Visual Communications and Image Processing 88*, T. R. Hsing, Ed., Cambridge, USA, November 1988, SPIE vol. 1001, pp. 616–624.

[22] J. Besag, "Spatial interaction and the statistical analysis of lattice systems", *Journal Royal Statistical Society B*, vol. 36, no. 2, pp. 192–236, 1974.

[23] J. Konrad, *Bayesian Estimation of Motion Fields from Image Sequences*, PhD thesis, McGill-University, Montreal, Canada, 1989.

[24] T. Aach and A. Kaup, "Partitioning of stereoscopic sequences by evaluation of stereo disparity and temporal change detection", in *Proceedings Picture Coding Symposium 91*, Tokyo, Japan, September 1991, pp. 443–446.

[25] J. Besag, "On the statistical analysis of dirty pictures", *Journal Royal Statistical Society B*, vol. 48, no. 3, pp. 259–302, 1986.

[26] F. Glazer, "Multilevel relaxation in low-level computer vision", in *Multiresolution Image Processing and Analysis*, A. Rosenfeld, Ed. 1983, pp. 312–330, Springer Verlag.

[27] D. Terzopoulos, "Image analysis using multigrid relaxation methods", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 2, pp. 129–139, 1986.

[28] D. Terzopoulos, "Multilevel computational processes for visual surface reconstruction", *Computer Vision, Graphics, and Image Processing*, vol. 24, pp. 52–96, 1983.

[29] S. S. Sinha and B. G. Schunck, "A two-stage algorithm for discontinuity-preserving surface reconstruction", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 1, pp. 36–55, 1992.

Til Aach was born in Cologne, Germany, on July 17th, 1961. He received the Dipl.-Ing. degree and the Doctoral degree, both in electrical engineering, from Aachen University of Technology (RWTH) in 1987 and 1993, respectively. In 1987, he joined the Institute for Communication Engineering, Aachen University of Technology, as a research assistant. His research interests are in image analysis and coding, statistical signal processing, and medical image processing.

T. Aach is a member of the Informationstechnische Gesellschaft (ITG/VDE).

André Kaup was born in Jugenheim, Germany, on June 4th, 1964. He received the Dipl.-Ing. degree in electrical engineering from Aachen University of Technology (RWTH) in 1989. Currently, he is with the Institute for Communication Engineering, Aachen University of Technology, where he is working as a research assistant towards his Doctoral degree. His current research interests are in statistical model-based image analysis and coding, and models for human vision.

A. Kaup is a member of the Informationstechnische Gesellschaft (ITG/VDE).