A non-local Stereo Similarity based on Collinear Groups

Nicolas Pugeault[1], Florentin Wörgötter[1] and Norbert Krüger[2]
[1]University of Stirling, Department of Psychology, Scotland, UK
[2]Aalborg University Esbjerg, Department of Computer Science, Denmark

October 26, 2003

Introduction

Vision, although widely accepted as the most powerful sensorial modality, faces the problem of an extremely high degree of vagueness and uncertainty in its low level processes such as edge detection, optic flow analysis and stereo estimation [1]. This arises from a number of factors. Some of them are associated with image acquisition and interpretation: owing to noise in the acquisition process along with the limited resolution of cameras, only rough estimates of semantic information (e.g., orientation) are possible. The severeness of these problems increases for higher semantic information, such as curvature or junction detection and interpretation. Furthermore, illumination variation heavily influences the measured grey level values and is hard to model analytically (see, e.g., [9]). Extracting information across image frames, e.g., in stereo and optic flow estimation, faces (in addition to the above mentioned problems) the correspondence and aperture problem which interfere in a fundamental and especially awkward way (see, e.g., [2, 10]).

However, by integrating information over context the human visual systems acquires visual representations which allows for actions with high precision and certainty within the 3D world even under rather uncontrolled conditions [16, 8]. The power of modality fusion arises from the huge number of intrinsic relations in visual data. The aim of the European project ECOVISION (see [4]) is to use such context knwoledge to achieve robust and more complete descriptions of the visual scene.

In this paper, we address a specific context in which aspects of 2D and 3D feature processing become combined. In human vision local visual entities become organised into more complex entities. This processes is usually called grouping (see, e.g., [19]). In computer vision such grouping processes are mostly treated within the image domain [18, 3]. Also in this paper, we start with a grouping process in the 2D image domain. However, this process becomes combined with stereo processing such that coherent 3D groups emerge. The constraint on which this combination is based is the following (see also figure 6):

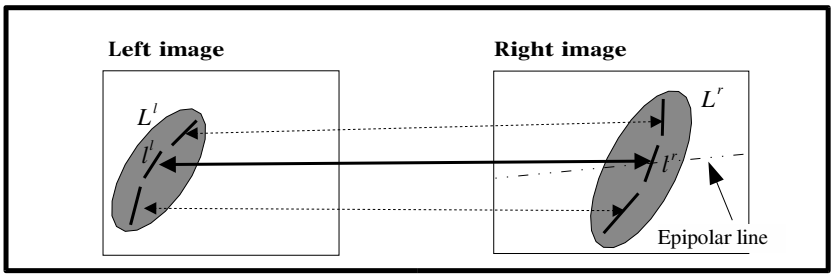

Figure 1: Stereo Collinearity Constraint

Stereo Collinearity Constraint: Primitives constituting a group in the left image have stereo correspondences in one group in the right image.

In this paper, we use this constraint to improve stereo processing. Stereo is necessarily ambiguous when based on local comparisons since the correspondence problem leads to mismatches. Using multiple modalities (such as colour or optic flow) improves but can not solve this problem (see [7, 12, 17]).

In this paper, we introduce an artificial visual system in which differerent processes are realized that support each other:

2D Feature Extraction: We have developed an image representation in form of 2D Primitives. These Primitives are multi-modal local descriptors that carry information about visual aspects such as orientation, contrast transition, colour and optic flow in a condensed way (see figure 2 and [13]).

2D Grouping: The 2D Primitives are local descriptors that become organised into higher entities in form of collinear groups. In the grouping process a linking structure is establised that makes use of a criterion that utilises collinearity as well as similarity in colour and contrast transition.

3D Feature Extraction by Stereo: We use the 2D-Primitives to find stereo correspondences. In this way we compute 3D Primitives from the 2D Primitives. The 3D Primitives carry information about 3D position and 3D orientation in addition to the information covered by the generating 2D Primitives.

Interaction of Stereo and Grouping: Finally, the group structures are used for improving stereo leading to coherent groups in 3D using the Stereo Collinearity Constraint.

The paper is structured as follows: In section 1 we briefly describe our processing of multi-modal Primitives. A more detailed description can be found in e.g., in [13]. The 2D grouping process is described in section 2. The multi-modal stereo is described in section 3 (further details can be found in [12, 17]) and the integration of grouping

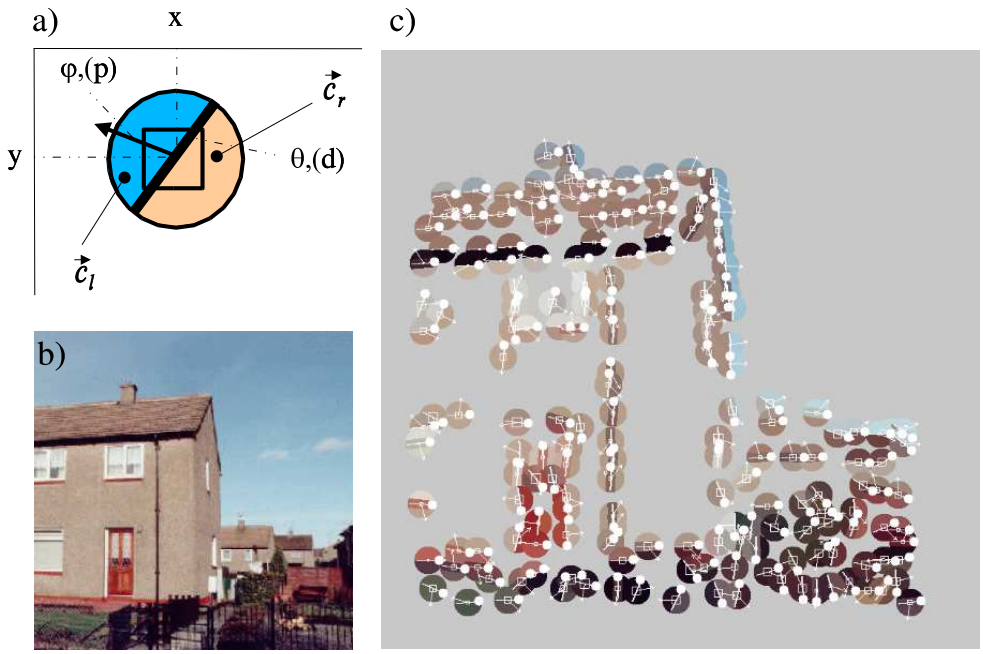

Figure 2: **Top left:** Schematic representation of a basic feature vector. Position is coded by (x, y), orientation by θ (or direction as d respectively), phase by φ (or p when associated with a direction), and color by $(\mathbf{c}_l, \mathbf{c}_r)$. **Bottom left:** Frame in an image. **Right:** Extracted feature vectors.

and stereo is described in secction 4. Results on artificial and real scenes are given in section 5.

1 Feature Processing

In this section we briefly describe the coding of information (orientation, phase and color) in terms of multi-modal Primitives.

Position, Orientation and Phase: We use a systematic mathematical description of geometric and structural information of grey level images based on the monogenic signal [6]. The monogenic signal performs a *split of identity*, i.e., it orthogonally divides the signal into energetic information (indicating the likelihood of the presence of a structure), its orientation θ and its structure (expressed in the phase φ). Features are extracted at energy maxima in local image patches where the position is parameterized by $\mathbf{x}$ (see figure 2).

The phase can be used to interpret the kind of contrast transition at this maximum [11], e.g., a phase of $\frac{\pi}{2}$ corresponds to a dark–bright edge, while a phase of 0 corresponds to a bright line on dark background. The continuum of contrast transition at an intrinsic one-dimensional signal patch can be expressed by the continuum of phases.

Color: The distribution of phases in natural images has been investigated in [14]. There

exist clear peaks at $\varphi = \pi/2$ and $\varphi = -\pi/2$ which show that edges (i.e., intrinsic 1-dimensional signals with odd symmetry) are the dominant one-dimensional structure in natural images while line structures (i.e., intrinsic 1-dimensional signals with even symmetry) are less dominant. Our model for an intrinsically one–dimensional signal patch (see figure 2) therefore describes edges.[1]

To integrate the modality color at intrinsically one–dimensional image structures we perform an averaging in the RGB color space over the left and right part ('left' and 'right' defined by the associated line segment) of the image patch (see figure 2).

We get two vectors $\mathbf{c}_l = (c_r^l, c_g^l, c_b^l)$ and $\mathbf{c}_r = (c_r^r, c_g^r, c_b^r)$, representing the red, green and blue values of the left and right side of the edge.

Therefore, the basic feature vector represented by our Primitives has the form

$$\mathbf{e} = (\mathbf{x}, \theta, \varphi, (\mathbf{c}_l, \mathbf{c}_r)).$$

2 Establishing Groups by a multi-modal Collinearity Criterion

We want to define groups of locally consistent Primitives in the image. We are interested in Primitives outlining major structures of the scenery, and subsequently of the images processed. We assume that any structure of the scene having a projective manifestation in the image has a representation involving a set of consistent Primitives (in the following called group). From this assumption follows naturally that Primitives showing inconsistency with their neighbourhood might be considered as ambigious information likely to be caused by erroneous feature extraction. In this section, we want to define the meaning of this *consistency* in the multi-modal space of the features.

In this work, we consider Primitives defining local oriented structures (e.g., lines and step edges). Therefore, we are looking for constellations defining global contours. Consistency between two Primitives is defined by two criterions: Collinearity and Modality Consistency (using the modalities colour and contrast transition). Inconsistency according to these two criterions indicates that the two Primitives are either expressions of independent structures or caused by the erroneous feature extraction process. In the following formulas we will consider a pair of Primitives $\mathbf{e}_1, \mathbf{e}_2$ such as $\mathbf{e}_2 \in N(\mathbf{e}_1)$, N being a large enough neighbourhood. We will consider the coordinate system centered in $\mathbf{e}_1$ and oriented so that $\theta(\mathbf{e}_1) = 0$. We want to define relationships between $\mathbf{e}_1$ and $\mathbf{e}_2$ defining possible structures for $\mathbf{e}_1$ and we code them as links $l(\mathbf{e}_1, \mathbf{e}_2)$ between them. We associate a confidence $c[l(\mathbf{e}_1, \mathbf{e}_2)]$ to a link which is an estimate of the probability for the two primitives to be part of the same structure.

2.1 Collinearity Criterion

Our collinearity criterion is based on two factors: Proximity and good continuation.

[1] Although there is significantly more edge like structures than line like structures in natural images we can also make use of an extra line model to describe intrinsically one-dimensional image patches with phase close to 0 or π. The introduction of this model makes only small difference for stereo matching (but is important in other contexts). We neglect this issue here.

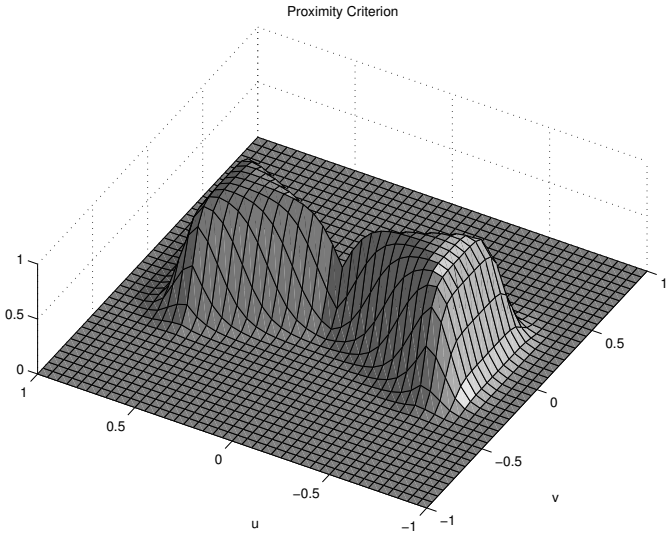

Figure 3: Proximity Criterion: Surface plot of the proximity function with the position of the second Primitive relative to the first

2.1.1 Proximity

Our proximity criterion take evaluate how given the position of the primitive $\mathbf{e}_2$ relatively to the primitive $\mathbf{e}_1$ a link $l(\mathbf{e}_1, \mathbf{e}_2)$ is likely to exist. The idea here is that the closer the second primitive is to the first, the closer it has to be to the line defined by the orientation of $\mathbf{e}_1$: parallel segments cannot be collinear for example. Also at this very local level we only want to consider low curvatures between the two primitives. To take these aspects into account we define a distance function between two Primitives by

$$C_{position}(\mathbf{e}_1, \mathbf{e}_2) = \frac{1}{1 + e^{\lambda(|x| - max(|y|, 0.3))}} \frac{1}{1 + e^{\lambda(|y| - 0.7)}} \tag{1}$$

with $\lambda = 30$ being the steepness parameter. A distance of 1 in the axis means twice the size of the patch generating the Primitives, and zero meaning the generating image patches of the two Primitives are in contact or overlapping. Figure 3 displays the distance function.

2.1.2 Good Continuation Criterion

If we consider the two modalities $\mathbf{e}_1$ and $\mathbf{e}_2$, the continuity in terms of orientation can be defined as a minimal curve joining $\mathbf{e}_1$ and $\mathbf{e}_2$. This curve ideally joins the positions A and B and is tangent to the orientation of $\mathbf{e}_1$ and $\mathbf{e}_2$ in those points.

In the following we consider the coordinate system O, u, v (see also figure 4) such as:

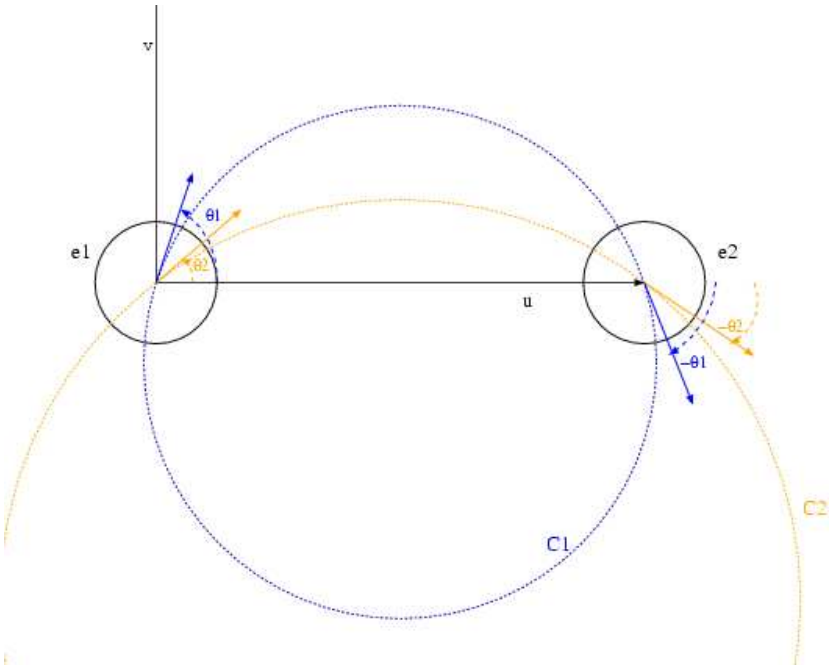

Figure 4: Good Continuation criterion: here we see that we can define an unique circle from the positions of $\mathbf{e}_1$ and $\mathbf{e}_2$ and the orientation of $\mathbf{e}_1$. This circle gives us an estimate for the orientation of $\mathbf{e}_2$

- O being the location of the first Primitive $\mathbf{e}_1$,
- u the vector from $\mathbf{e}_1$ to $\mathbf{e}_2$
- v normal to u

The axis are normalized so that a distance of 1 is the distance between $\mathbf{e}_1$ and $\mathbf{e}_2$ in the image.

Consequently, the position of $\mathbf{e}_1$ is defined by the vector $p(0) = (0, 0)$ and $\mathbf{e}_2$ by $p(1) = (1, 0)$ We can define a unique circle from the positions of $\mathbf{e}_1$ and $\mathbf{e}_2$ and the orientation θ_1 of $\mathbf{e}_1$. This circle gives us an estimate for the orientation of $\mathbf{e}_2$ (see figure 4). Actually it mus hold $\theta_2 = -\theta_1$. An estimation of the likelihood of the curve defined by the two Primitives is then defined using the difference between this estimated orientation and the measured one (see figre 4).

$$p'(1) = -p'(0) = -tan(\theta_1) \tag{2}$$

$$Coll(\mathbf{e}_1, \mathbf{e}_2) = |tan(\theta_2) + tan(\theta_1)| \tag{3}$$

2.2 Modality Continuity Criterion

The consistency over the color and phase modalities is calculated using the similarity functions for phase $sim_\phi(\mathbf{e}_1, \mathbf{e}_2)$ and colour $sim_C(\mathbf{e}_1, \mathbf{e}_2)$ already used in [12, 17]. Here we assume that modalities are continuous over a given 3D feature. Consequently, they should be continuous over their manifestation in the image. We say that a link $l(\mathbf{e}_1, \mathbf{e}_2)$ exists when the similarities in the modalities are high enough. Consequently we define an estimate for the consistency of the pair $c[l(\mathbf{e}_1, \mathbf{e}_2)]$ by

$$c[l(\mathbf{e}_1, \mathbf{e}_2)] = Coll(\mathbf{e}_1, \mathbf{e}_2) \cdot (sim_\phi(\mathbf{e}_1, \mathbf{e}_2) + sim_C(\mathbf{e}_1, \mathbf{e}_2)) \,. \tag{4}$$

An example of the links confidences for our test sequences can be seen in figure 5.

3 Multi-modal stereo

To create 3D information from the 2D Primitives by stereo we need to match Primitives in the left and right image. In [12, 17] we have derived a matching function makes use of information in all modalities. A pair or Prmitives $(\mathbf{e}^l, s(\mathbf{e}^l))$ represents a Primitive in the left image and its correspondent in the right image ($s(\mathbf{e}^l)$ being the matched Primitive in the right image). From such a correspondence we can compute a 3D Primitive $\mathbf{E}$ by a reconstruction function R (see, e.g., [5]):

$$\mathbf{E} = R(\mathbf{e^l}, s(\mathbf{e^l})) \tag{5}$$

Moreover, every Primitive has a list of potential stereo-correspondences containing all Primitives of the second image intersecting the epipolar line drawn from the first Primitive. The i-th entry of the list is denoted by $s_i(\mathbf{e^l})$. In [12, 17] only the best correspondence is used to generate the 3D-entity. The decision between several potential

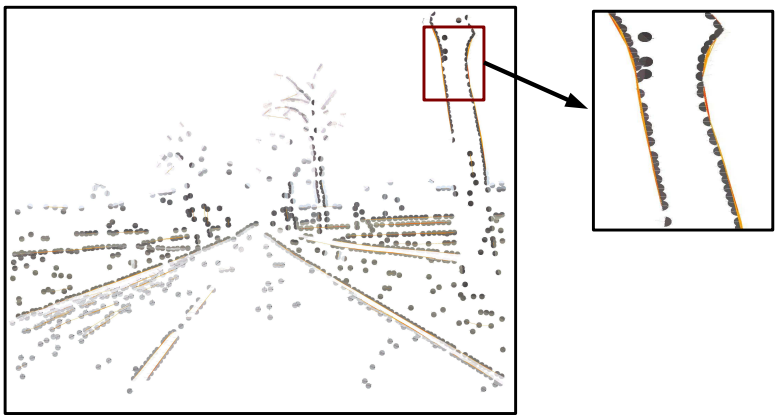

Figure 5: The potential links between the primitives are shown by the orange lines. The darker the line, the higher the confidence in the link.

matches is made comparing similarities in local modality measurements of both primitives. We will call this estimation of the quality of a match the *internal confidence* and note it $c[s(\mathbf{e})]$: it is all that can be estimated using the locally available information of the Primitive.

Stereo Matching based on the internal confidence is naturally ambiguous, for example repetitive structures may occur in a scene leading to similar Primitives for distinct scene elements. Also due to projective distortion between both images the actual similarity might be misleading: for example differences in orientation and colour can be expected in both images according to the different perspective views of the left and right image. This difference of course cannot be anticipated in a local way leading to sub-optimal similarity estimation. Consequently, the internal confidence on its own is a naturally inaccurate and ambiguous measure.

4 Combining Grouping and Stereo

In this paper, we want to improve the decision based of local information by taking into account the consistency over the Primitive's neighbourhood utilising the grouping process defined in section 2. The core idea is to compare how similar neighbourhood of the potential matches are to the neighbourhood of the original Primitive to define an *external confidence* in the match (written $c_{ext}[s(\mathbf{e})]$). The neighbourhood is here considered as the network of links associated to the Primitive.

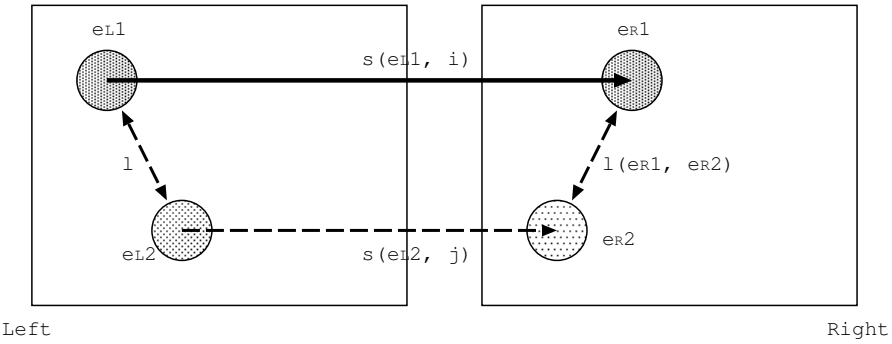

Figure 6: The BSCE criterion: Given a stereo correspondence $s_i(\mathbf{e}_1)$, the BSCE can be calculated for a primitive $\mathbf{e}_2$ in the neighbourhood, depending on $l(\mathbf{e}_1, \mathbf{e}_2)$, $s_j(\mathbf{e}_2)$, and $l'(s_i(\mathbf{e}_1), s_j(\mathbf{e}_2))$. The bold line represent the event we want to confirm, and the dashed lines the external events which, in conjonction, confirms it.

4.1 Stereo-Consistency Element

We considered that consistency in Primitives was not incidental but a consequence of the scene structure and therefore this consistency should be conserved by stereo (except, of course in case of occlusion). We want to define a stereo correspondence mechanism handling this external confidence based on the following principles:

Postponement of early hard Decision: Differing from (5) we want postpone the decision of a succesful match and allow for multiple correspondences leading to mutiple potential 3D matches. The final decision is done after the grouping process considering the Stereo Collinearity Constraint.

Uniqueness leads to Competition: As stereo correspondences are mutually exclusive competition needs to be included in any correction/adaptation process.

Weighting according to Group Consistency: Over one Primitive neighbourhood, the relative weight of the stereo correspondence of a neighbour is proportional to the consistency of the Primitive with this neighbour (i.e. to the link confidence).

Weighting according to Stereo Consistency: The influence of a Primitive over its neighbours is proportional to the confidence in its stereo-correspondences (consequently a Primitive with only poor stereo correspondences will do little to help stereo decisions).

We now define the *minimal* stereo event involving a primitive neighbourhood: Given two Primitives $\mathbf{e}_1^L$ and $\mathbf{e}_2^L$ in the left frame such as a link $l(\mathbf{e}_1^L, \mathbf{e}_2^L)$ can be defined between them, if we consider the hypothesis that $s_i(\mathbf{e}_1^L)$ is the correct stereo-correspondence for $\mathbf{e}_1^L$ in the right image:

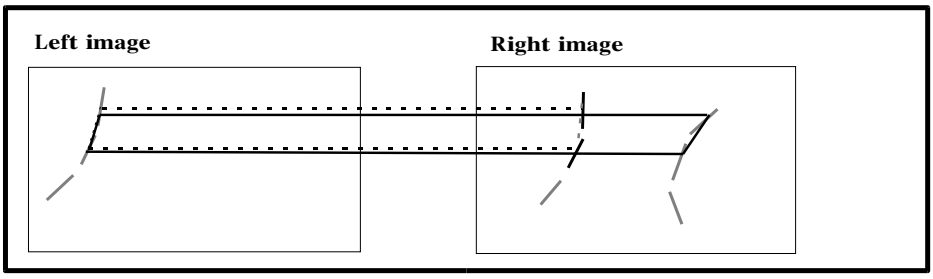

Figure 7: Since one Primitive can have multiple matches it can be verified by multiple BSCEs. The dotted and bold structures each represent one BSCE.

if there exists a link $l(s_i(\mathbf{e}_1^L), s(\mathbf{e}_2^L))$ between this stereo-correspondence and the public stereo-correspondence $s(\mathbf{e}_2^L)$ of the second primitive $\mathbf{e}_2^L$

then the hypothesis $s(\mathbf{e}_1^L)$ is confirmed

else this hypothesis is contradicted.

We call this the *Basic Stereo Consistency Event* (BSCE) (see figure 6).

4.2 BSCE confidence

We want to associate a confidence to the BSCE event. Here we are not working with certainties, but with potential links and stereo-correspondences. Consequently we want a continuous formulation of the BSCE, giving us a confidence in its realization. We propose in this section to draw this from the previous confidences of the basic events involved.

First, we define a set of function that are used at different places:

$$\begin{array}{lcll} f^g(a_1,\ldots,a_n) & = & (a_1 \cdot \ldots \cdot a_n)^{\frac{1}{n}} & \text{Geometric Mean} \\ f^a(a_1,\ldots,a_n) & = & \frac{a_1+\ldots+a_n}{n} & \text{Arithmetic Mean} \end{array}$$

The geometric mean represents a harder connection between events than the arithmetic mean. The multiplication works like a "logical and" ($\wedge$) while the arithmetic mean is a softer connection. We apply the arithmetic mean when different cues cooperate "democratically" while the geometric mean is used when the non-occurence of one event supresses all others.

Now the confidence associated to a BSCE can be estimated from the known confidences as follows:

$$c[BSCE_i(\mathbf{e}_1^L, \mathbf{e}_2^L)] = f^g\left(c[l(\mathbf{e}_1^L, \mathbf{e}_2^L)], c[s(\mathbf{e}_2^L)], c[l(s_i(\mathbf{e}_1^L), s(\mathbf{e}_2^L))]\right) \qquad (6)$$

where f^g represents the geometric mean.

4.3 Neighbourhood Consistency Confidence

Equation (6) gives us how a Primitive stereo correspondence is consistent with our beliefs on another Primitive stereo properties. We now want to estimate how this correspondence is consistent with the *whole neighbourhood* of the Primitive. Now if we consider a primitive $\mathbf{e}_1^L$ and an associated stereo-correspondence $s_i(\mathbf{e}_1^L)$, we can integrate this BSCE confidence over the neighbourhood of the primitive ($N_{\mathbf{e}_1^L}$). We call this confidence the *external confidence* in the stereo-correspondence:

$$c_{ext}[s_i(\mathbf{e}_1^L)] = \frac{1}{|N_{\mathbf{e}_1^L}|} \sum_{\mathbf{e}_k^L \in N_{\mathbf{e}_1^L}} c[BSCE_i(\mathbf{e}_1^L, \mathbf{e}_k^L)] \tag{7}$$

This gives us a confidence on how consistent is a stereo-correspondence with the stereo of the Primitive neighbourhood. Note that equation 7 represents a non–local stereo similarity!

4.4 Outlier Removal Process

In the outlier removal process we are after the reliable matches (i.e., we want to eliminate possibly false matches). The outlier removal process can be used where a small number of reliable features is used to compute the motion between frames (in this case we need reliable 3D-2D matches).

Our actual system rank the potential correspondences of a Primitive depending on their similarity (over all modalities) with this primitive, and the best one (or public one) is assumed to be the correct correspondence. We propose here to threshold the external confidence of those potential correspondences in order to remove those in contradiction with their neighbourhood current assumptions (i.e. neighbours public correspondences). We expect this way to remove wrong correspondences, otherwise impossible to discern from correct ones using local modalities. Figures 8 and 9 show the result of the outlier removal process.

5 Results

We have applied this outlier removal process to two stereo sequences. The first one (fig. 8) is a simple artificial scene generated using openGL. The second scene has been recorded near Lippstadt (Germany) from a pair of calibrated cameras fixed to a car (with the cooperation of HELLA). This second scene (fig. 9) represent more accurately the standard conditions in which a natural system has to operate (low saturation, highly textured surfaces, etc...). Both figures show the left and right images on the top row. On the middle row, the images show the primitives extracted. The red lines reach to the position of their current public correspondence in the right image. Those pairs (from the public correspondences of each primitive) are used to reconstruct 3D entities. The lower figure show a reprojection of those 3D entities on the horizontal plane (the horizontal axis is the Z axis, and the vertical axis is the X axis here). The left two pictures show the original public correspondences and reconstruction using

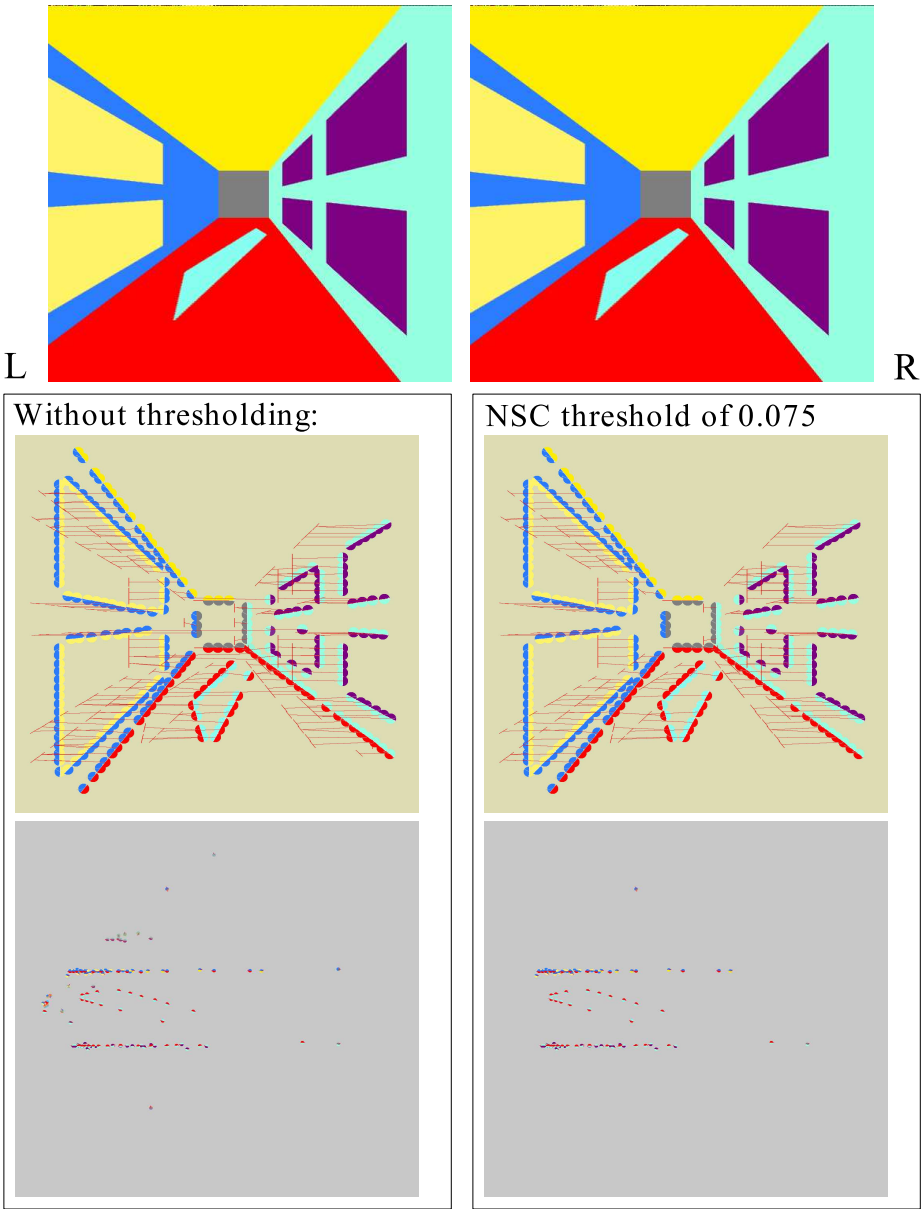

Figure 8: We apply our external confidence thresholding to this artificial scene. The left two images represent the results without thresholding, and the right ones with thresholding. In both case, the middle image show the primitives extracted by our program, and the lines reach to the position of their current public correspondence in the right image. The lower one show a orthographic reprojection of the reconstructed 3D entities (from those public stereo pairs). This shows the XZ (horizontal) plane.

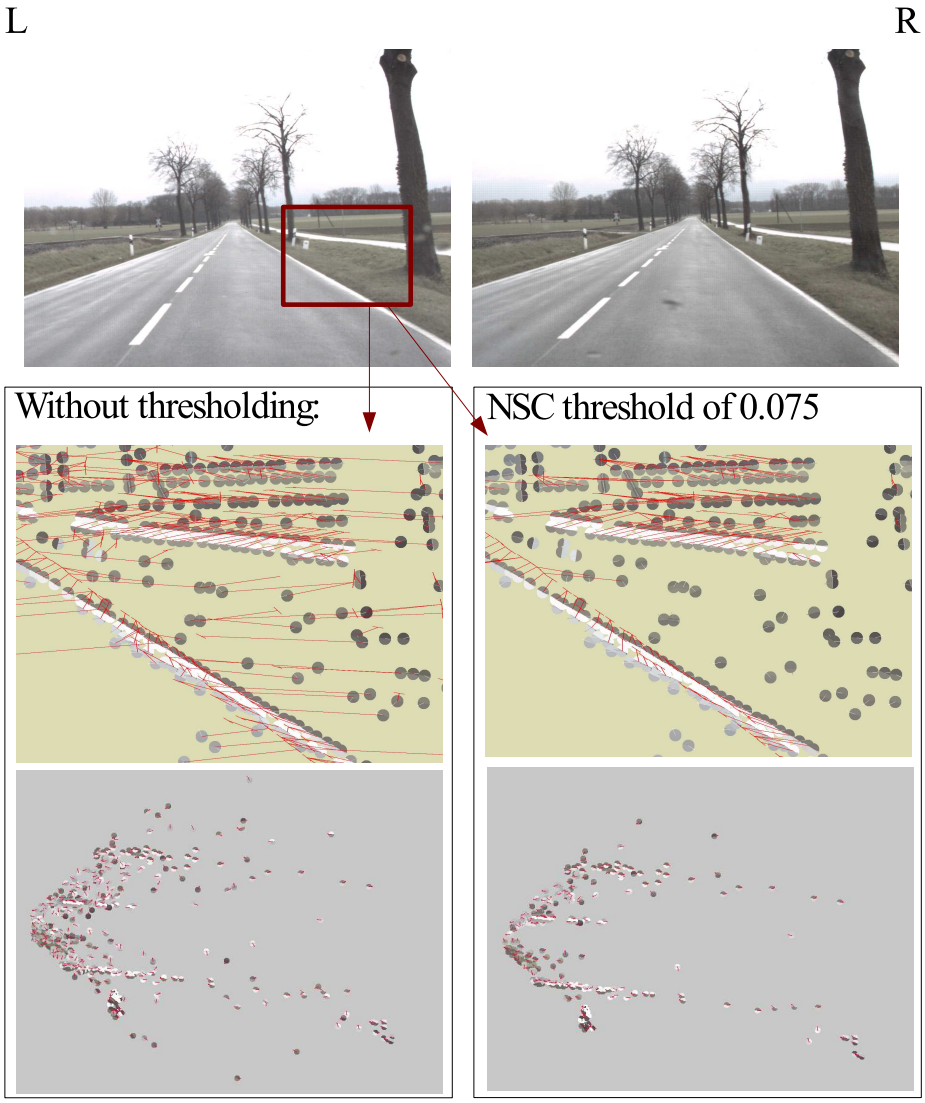

Figure 9: This figure show the same results as 8 with a natural stereo scene this time. The middle images show here a zoom in of the primitives. Here the correspondences of the primitives created by the texture are being removed, while consistent lines are being preserved.

only internal confidence. The two pictures on the right of the figures show the public correspondences and the resulting 3D reconstructed entities after a thresholding of the correspondences over their external confidence (the threshold is of 0.075 in both cases).

On the figure 8 most of the wrong correspondences are being removed through this process. More interestingly on the figure 9, showing the difficult natural scene, a considerable amount of noise is being removed. In the magnified view of the correspondences, we can see that most random correspondences from primitives extracted from texture artifact are being removed, while consistent correspondences are preserved. Note that in both case this improvement is gained *only by thresholding the external confidence*, and without any additional thresholding on the actual similarity of the Primitives.

6 Summary and Outlook

In this paper we have introduced a non-local stereo similarity function that makes use of a grouping process. We could show that a good matching could be achieved using a *non-loacl criterion*. We coild further show that a large number of outliers could be removed by using the contextual information coded in the grouping process.

The combination of grouping processes processes in the 3D and time domain has been recognised as a field which has only been addressed by a few scientists. on (see, e,g., [18]). However, we think that grouping can have a significant impact on the modelling of processes such as stereo, structure from motion and motion estimation. We see the combination of grouping processes with low–level feature processing (such as stereo) as we as higher cognitive functions such as object categorisation as an emergent field that can overcome weaknesses of today's artificial visual sysems (for an overview see [15]).

References

[1] Y. Aloimonos and D. Shulman. *Integration of Visual Modules — An extension of the Marr Paradigm*. Academic Press, London, 1989.

[2] N. Ayache. *Stereovision and Sensor Fusion*. MIT Press, 1990.

[3] K.L. Boyer and S. Sarkar. Perceptual organization in computer vision: Status, challenges, and potential. *Special Issue on Perceptual Organization in Computer Vision, October*, 76(1):1–5, 1999.

[4] ECOVISION. Artificial visual systems based on early-cognitive cortical processing (EU–Project). *http://www.pspc.dibe.unige.it/ecovision/project.html*, 2003.

[5] O.D. Faugeras. *Three–Dimensional Computer Vision*. MIT Press, 1993.

[6] M. Felsberg and G. Sommer. The monogenic signal. *IEEE Transactions on Signal Processing*, 49(12):3136–3144, December 2001.

[7] P.B. Hibbard, M.F. Bradshaw, and R.A. Eagle. Cue combination in the motion correspondence problem. *Proceedimgs of the Royal Society London B*, 267:1369–1374, 2000.

[8] D.D. Hoffman, editor. *Visual Intelligence: How we create what we see*. W.W. Norton and Company, 1980.

[9] K. Ikeuchi and B.K.P. Horn. Numerical shape from shading and occluding boundaries. *Artificial Intelligence*, 17:141–184, 1981.

[10] R. Klette, K. Schlüns, and A. Koschan. *Computer Vision - Three-Dimensional Data from Images*. Springer, 1998.

[11] P. Kovesi. Image features from phase congruency. *Videre: Journal of Computer Vision Research*, 1(3):1–26, 1999.

[12] N. Krüger, M. Felsberg, C. Gebken, and M. Pörksen. An explicit and compact coding of geometric and structural information applied to stereo processing. *Proceedings of the workshop 'Vision, Modeling and VISUALIZATION 2002'*, 2002.

[13] N. Krüger, M. Lappe, and F. Wörgötter. Biologically motivated multi-modal processing of visual primitives. *AISB*, 1(4), 2003.

[14] N. Krüger and F. Wörgötter. Multi modal estimation of collinearity and parallelism in natural image sequences. *Network: Computation in Neural Systems*, 13:553–576, 2002.

[15] N. Krüger and F. Wörgötter. Statistical and deterministic regularities: Utilisation of motion and grouping in biological and artificial visual systems. *Advances in Imaging and Electron Physics*, to appear.

[16] W.A. Phillips and W. Singer. In search of common foundations for cortical processing. *Behavioral and Brain Sciences*, 20(4):657–682, 1997.

[17] N. Pugeault and N. Krüger. Multi–modal matching applied to stereo. *Proceedings of the BMVC 2003*, 2003.

[18] S. Sarkar and K.L. Boyer. *Computing Perceptual Organization in Computer Vision*. World Scientific, 1994.

[19] R.J. Watt and W.A. Phillips. The function of dynamic grouping in vision. *Trends in Cognitive Sciences*, 4(12):447–154, 2000.