A SIGNAL-SYMBOL LOOP MECHANISM FOR ENHANCED EDGE EXTRACTION

Preparation of Camera-Ready Contributions to INSTICC Proceedings

Sinan Kalkan, Florentin Wörgötter
Bernstein Center for Computational Neuroscience, Univ. of Göttingen, Germany
{sinan,worgott}@bccn-goettingen.de

Shi Yan, Volker Krüger
Medialogy Lab, Aalborg Univ. Copenhagen, Denmark
syan06@imi.aau.dk, vok@media.aau.dk

Norbert Krüger
Cognitive Vision Lab, Univ. of Southern Denmark, Denmark
norbert@mip.sdu.dk

Keywords: Feedback Mechanisms, Edge Extraction, Rigid Body Motion, Signal–Symbol Loop

Abstract: The transition to symbolic information from images involves in general the loss or misclassification of information. One way to deal with this missing or wrong information is to get feedback from concrete *hypotheses* derived at a symbolic level to the sub-symbolic (signal) stage to amplify weak information or correct misclassifications. This paper proposes such a feedback mechanism between the symbolic level and the signal level, which we call *signal symbol loop*. We apply this framework for the detection of low contrast edges making use of predictions based on Rigid Body Motion. Once the Rigid Body Motion is known, the location and the properties of edges at a later frame can be predicted. We use these predictions as feedback to the signal level at a later frame to improve the detection of low contrast edges. We demonstrate our mechanism on a real example, and evaluate the results using an artificial scene, where the ground truth data is available.

1 INTRODUCTION

Processing in most artificial vision systems as well as in the human visual system starts with the extraction of information based on linear and non-linear filtering operations (figure 1) by which, *e.g.*, local orientation, magnitude, and phase become computed. We call this level of processing '*signal*-level' since the original signal is usually reconstructible from it; *i.e.*, the signal-level information is pixel-wise, continuous and complete.

In a next step, we extract discrete descriptors for line structures using the method of (Krüger et al., 2004). We call this level '*symbol*-level' since at this stage the semantic information represented in single pixel values is made explicit. Symbolic information is sparse, condensed and semantically rich, and usually, the original signal is not fully reconstructible from it.

Inclusion of contextual information requires the exchange of information over large spatial or temporal distances (in case of, e.g., large object motions or saccades) and even the use of world knowledge stored in long term memory (as for example in the Dalmatian dog illusion (Gregory, 1970)[1]). Such exchange of information can only be formulated sub-optimally on the signal-level in a pixel-wise representation since the number of pairwise relations would simply become too large or the amount of computer memory required would exceed reasonable bounds. The advantage of a symbolic level is that reasoning over spatial and temporal changes as well as interaction with the world knowledge stored in the memory becomes much easier. In this paper, we introduce a framework of, so called, 'signal-symbol loops' and apply it in the context of edge extraction.

The transition to the symbolic level requires the transformation of information at the pixel-wise and continuous signal level to a discrete and condensed symbolic level. This usually requires the use of thresholds. Binary decisions involving such a thresholding usually results in either a loss of information below the threshold or in the extraction of false positives caused by signal noise (see figure 2). In the case of finding line segments, for example, a threshold is introduced to determine *contrast sensitivity*.

[1] The illusion is also available online at `http://www.michaelbach.de/ot/cog_dalmatian/index.html`

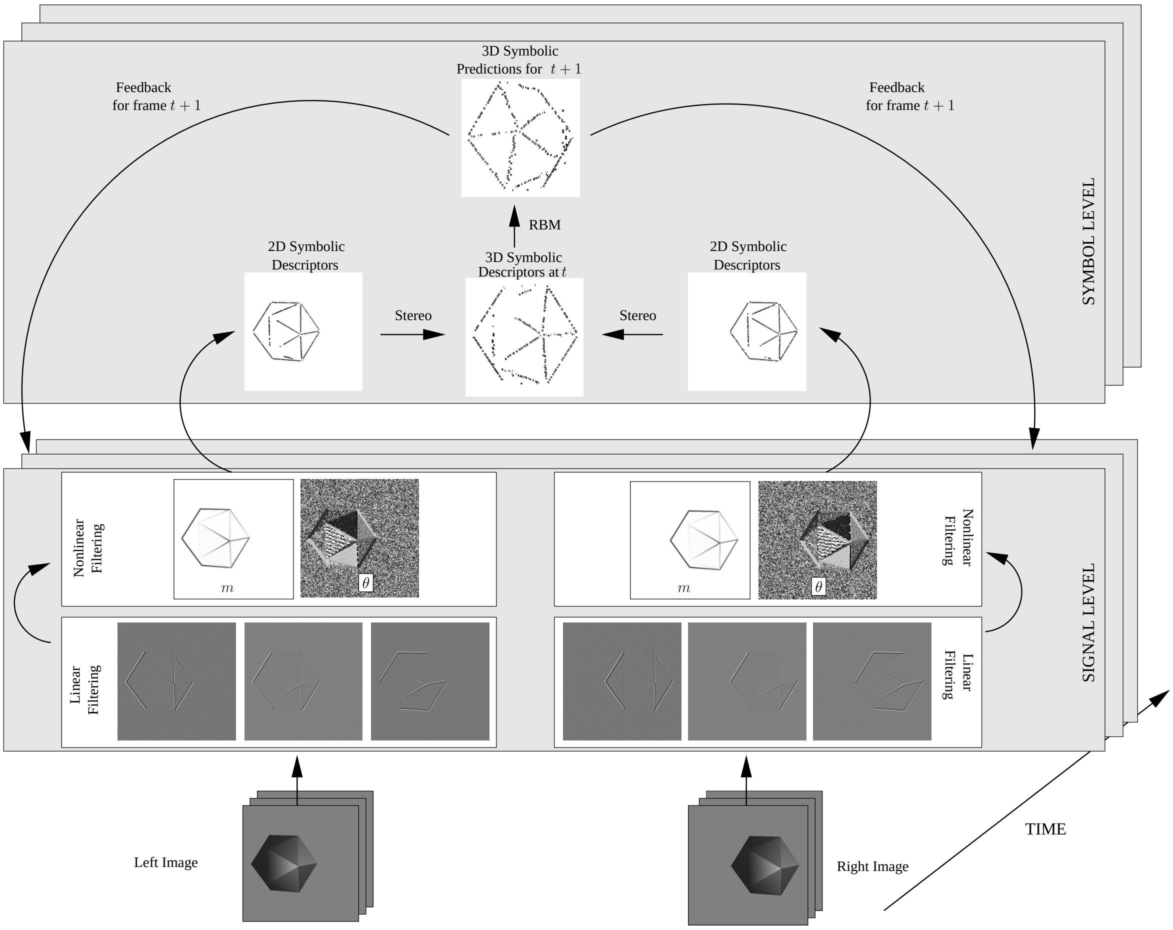


Figure 1: A rough outline of the signal-symbol loop mechanism, which is proposed in this paper. The linear filtering is achieved by Gabor wavelets (only real components of the three out of eight responses are shown). The non-linear filtering level contains the magnitude *m* and the orientation θ information. From the signal-level information, 2D symbolic edge descriptors are extracted. These descriptors are then matched to the other camera view to reconstruct 3D symbolic edge descriptors. The known RBM is used to estimate the 3D symbolic descriptors at a later frame $t+1$, whose projections to the respective images at time $t+1$ then provide the feedback to the filter processing layer. Note that the predicted 3D descriptors at frame $t+1$ are shown from a different perspective, and therefore, are not as smooth as the 3D primitives at frame *t*.

Using a high threshold (*i.e.*, low contrast sensitivity) produces reliable (*i.e.*, true positive) but (most of the time) incomplete set of line segments (figure 2). Using a low threshold (*i.e.*, high contrast sensitivity), on the other hand, can produce a more complete set of line segments, which usually include also noisy information (figure 2). This dilemma between *incomplete-but-reliable* versus *complete-but-noisy* is faced by all computer vision algorithms which require some thresholding. By local processing alone relevant information can not be distinguished from information caused by, *e.g.*, signal noise or other sources of ambiguity.

One way to gain the information lost during the transition to the symbolic level is to review the signal based on concrete hypotheses generated by reasoning on the symbolic level being *fed back* to the signal level to *amplify* the weak but consistent information. We call this feedback mechanism '*signal-symbol loop*' (see also (Krüger, 2005)).

To make information at the symbolic level comparable to the signal, it is required to transform the symbolic information back in a form that makes it comparable to the signal level. This transformation can be regarded as taking the inverse of a symbolic description, and therefore it is called the *feedback function* in the rest of the paper. This feedback function can be considered as the inverse of a symbol since it transform the symbolic information back to the signal-level information.

This paper proposes a concrete signal-symbol loop mechanism to improve the extraction of low-

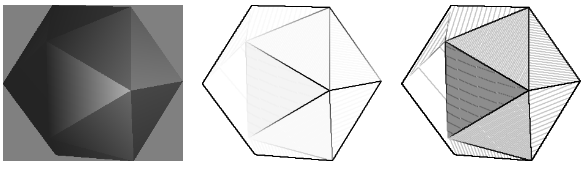


Figure 2: **(a)** An artificial image with low-contrast edges. **(b)** The result of the Sobel operator (Nixon and Aguado, 2002) with a high threshold. **(c)** The result of the Sobel operator with a low threshold (in order to extract the weak edges), which produces unwanted edges due to the shading ((c) is scaled independently for the sake of better visibility).

contrast edges by making use of motion information, namely, the change of a symbolic local edge descriptor under a Rigid Body Motion (RBM). In our paper, the change of position and orientation of this descriptor under an RBM can be formulated explicitly: After estimating the position of a 3D edge descriptor at a later frame, the image projection of the estimated 3D descriptor provides feedback to the filter processing level. The feedback information states that there must be an edge descriptor with certain properties at a certain position. The filter processing level then enhances the information at a position if the feedback is consistent with the original image information. The rough outline of the mechanism that we propose is given in figure 1.

The approach we introduce here is related to 'adaptive thresholding' approaches which are for example used in the area of image segmentation. These can also recover low-contrast edges by adjusting the threshold. This adjustment, however, is based on the *local* distribution of image intensities (see, *e.g.*, (Gonzales and Woods, 1992)). Our approach differs from adaptive thresholding since it makes use of symbolic information that facilitates a more global and also a more directed mechanism rather than local intensity distribution. Moreover, as we discuss at the end of the paper, the novelty of the current paper is in the proposal of a symbol-to-signal feedback mechanism that can be applied also in other contexts.

The idea of using of feedback in vision systems is not new (Aloimonos and Shulman, 1989; Angelucci et al., 2002; Galuske et al., 2002; Bullier, 2001). For computational models the interested reader is directed for example to (Bayerl and Neumann, 2007) for motion disambiguation or (Bullier, 2001) for modelling at the neuronal level for long-range information exchange between neurons. Our work is different from the above mentioned works in that we introduce a feedback mechanism between different layers of processing, *i.e.*, the signal-level and the symbol-level, and we apply it in a different context.

The paper is organized as follows: In section 2, we introduce the symbolic edge descriptors and the concept of RBM that are utilized in this paper. Section 3 describes our feedback mechanism. In section 4, we present and discuss the results, and the paper is concluded in section 5.

2 SYMBOLIC DESCRIPTORS AND PREDICTIONS

In this section, we give a brief description of the image descriptors that we use to represent local scene information at the symbolic level (section 2.1). These descriptors represent local image information in a condensed way and by that transform the local signal information to a symbolic level. In section 2.2, we briefly comment on Rigid Body Motion which we use as the underlying regularity of predictions on the symbolic level.

2.1 Multi-modal Primitives

The concept of multi-modal primitives has been first introduced in (Krüger et al., 2004). These primitives are local multi-modal scene descriptors, which are motivated by the hyper-columnar structures in V1 (Hubel and Wiesel, 1969).

In its current state, primitives can be edge-like or homogeneous and carry 2D or 3D information. For the current paper, only edge-like primitives are relevant. An edge-like 2D primitive (figure 3(a)) is defined as:

$$\pi = (x, \theta, \omega, (\mathbf{c}_l, \mathbf{c}_m, \mathbf{c}_r)), \tag{1}$$

where x is the image position of the primitive; θ is the 2D orientation; ω represents the local phase, the color is coded as three vectors $(\mathbf{c}_l, \mathbf{c}_m, \mathbf{c}_r)$, corresponding to the left $(\mathbf{c}_l)$, the middle $(\mathbf{c}_m)$ and the right side $(\mathbf{c}_r)$ of the primitive. See (Krüger et al., 2004) for more information about these modalities and their extraction. Figure 4 shows the extracted primitives for an example scene.

A primitive π is a 2D descriptor which can be used to find correspondences in a stereo framework to create 3D primitives (as introduced in (Krüger et al., 2004)) which have the following formulation:

$$\Pi = (X, \Theta, \Omega, (\mathbf{c}_l, \mathbf{c}_m, \mathbf{c}_r)), \tag{2}$$

where X is the 3D position; Θ is the 3D orientation. Appearance based information is coded by generalising local phase and color of the two corresponding

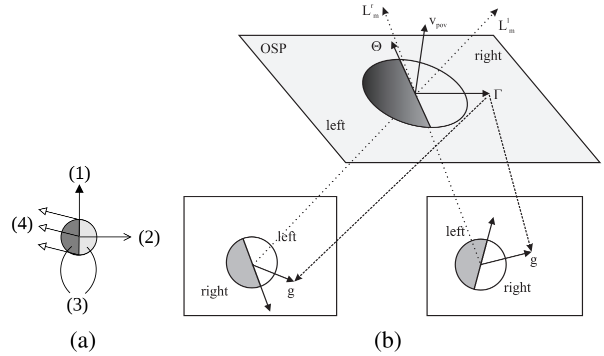


Figure 3: **(a)** An edge-like primitive: 1) represents the orientation of the primitive, 2) the phase, 3) the color and 4) the optic flow. **(b)** Two corresponding 2D edge primitives can reconstruct a 3D primitive.

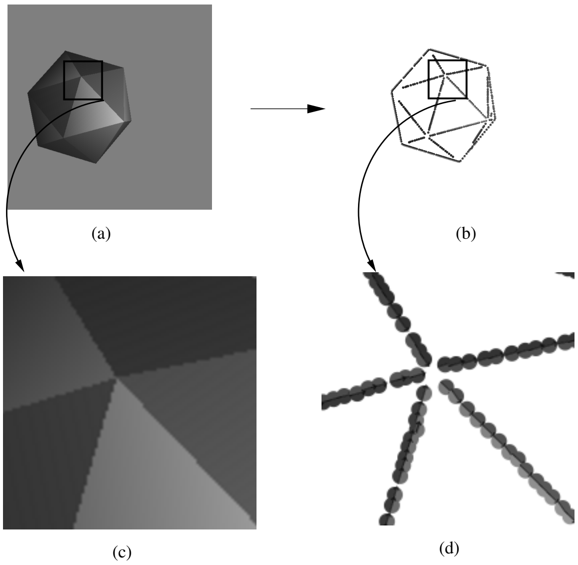


Figure 4: Extracted *edge* primitives (b) for the example image in (a). Extracted primitives for the region of interest in (c) is shown in (d).

2D primitives. The reconstruction of a 3D primitive from two corresponding 2D primitives is exemplified in Figure 3(b).

Knowledge of the camera parameters allows defining a projection relation $\mathcal{P}$ from a 3D primitive Π to an image, which produces a 2D primitive $\hat{\pi}$:

$$\hat{\pi} = \mathcal{P}(\Pi). \tag{3}$$

The projection $\hat{\pi}$ of a 3D primitive Π is used in section 3 for computing the feedback of a prediction.

2.2 Rigid Body Motion (RBM)

A Rigid Body Motion describes the motion (*i.e.*, translation and rotation) of rigid objects; *i.e.*, objects where the distance between any two particles on the object remains the same throughout the motion.

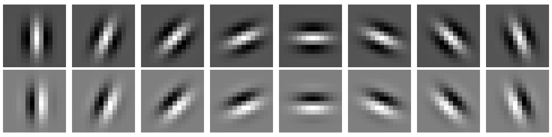


Figure 5: Real (first row) and imaginary (second row) parts of eight orientation Gabor wavelets.

A RBM associates a 3D entity e^t in the first frame to another entity $e^{t+\Delta t}$ in the second frame:

$$e^{t+\Delta t} = RBM(e^t). \tag{4}$$

Application of equation 4 requires computation of rotation and translation, which can be achieved by finding correspondences between 3D entities e^t and $e^{t+\Delta t}$ (see, *e.g.*, (Faugeras, 1993)).

Knowledge of the RBM allows estimation of the 3D entities, in our case the primitives, at a later frame:

$$\hat{\Pi}_i^{t+\Delta t} = RBM_{t\to t+\Delta t}(\Pi_i^t). \tag{5}$$

In this paper, the ground truth RBM is known either because the scene is generated using OpenGL, or because the object is rotated with a robot arm whose motion is known. See (Faugeras, 1993) for more information about RBM and RBM estimation methods.

3 FORMALIZATION OF THE SIGNAL-SYMBOL LOOP

The RBM predicts a 3D primitive at a later frame. This prediction is formulated at the symbolic level since it uses the 3D primitives. The projection of this primitive from the symbolic level into the image (using the projection relation defined in equation 3) provides a position and an orientation feedback to the filtering operations (*i.e.*, the signal level). At the filter-processing level, this feedback at discrete positions is combined with the extracted filter responses.

At the signal level, we use complex Gabor wavelets as a basic filtering operation (Lee, 1996). The Complex Gabor wavelet response G is computed on eight different orientations; *i.e.*, $G(x,y,c_i)$ for $i \in [1,8]$ (figure 5). The feedback of a prediction with image coordinate (x_0,y_0) and orientation θ_0 (falling into channel c_0)[2] is distributed over the Gabor channels

[2] The channel c_i that an orientation $\theta \in [0,\pi)$ corresponds to is computed using $i = round(N \cdot \theta/\pi)$ where $N = 8$ is the total number of channels.

using the following Gaussian Feedback Function:

$$\begin{aligned} F(x,y,c_i) &= \frac{1}{C}\ exp\Big(-\frac{1}{2}\Big\{ \\ &\quad \frac{[(x-x_0)\cos\theta_0+(y-y_0)\sin\theta_0]^2}{\sigma_x}+ \\ &\quad \frac{[-(x-x_0)\sin\theta_0+(y-y_0)\cos\theta_0]^2}{\sigma_y}+ \\ &\quad \frac{(c_i-c_0)^2}{\sigma_c}\Big\}\Big), \qquad (6) \end{aligned}$$

where C is a normalization constant computed using:

$$C=\frac{1}{(2\pi)^{1/2}(\sigma_x^2+\sigma_y^2+\sigma_\theta^2)}, \qquad (7)$$

where we empirically set $\sigma_x = 4, \sigma_y = 1, \sigma_\theta = 1$. The Gaussian Feedback Function in equation 6 is an essential part of the signal-symbol loop proposed in this paper since it distributes the incomplete, condensed and discrete symbolic information in a 2D primitive $\hat{\pi} = \mathcal{P}(RBM(\pi))$ to the complete, continuous and pixel-wise signal-level information: *i.e.*, $F(\hat{\pi}) = F(x,y,c_i)$ for $i = 1,..,8$.

The original Gabor responses and the feedback $F(x,y,c_i)$ from the symbolic level, *i.e.*, RBM estimation, are combined into a modified Gabor function $\hat{G}(x,y,c_i)$ as follows:

$$\begin{aligned} \hat{G}^R(x,y,c_i) &= G^R(x,y,c_i)+w\cdot F(x,y,c_i), \quad (8) \\ \hat{G}^I(x,y,c_i) &= G^I(x,y,c_i)+w\cdot F(x,y,c_i). \quad (9) \end{aligned}$$

where G^R and G^I are the complex and the imaginary parts of the respective orientation channels. We determine the weight w based on the consistency of the predicted orientation (*i.e.*, the orientation of the 2D projection of the predicted 3D primitive) with the extracted Gabor responses as follows:

$$w=\left[1-\frac{1}{N\cdot\pi/2}\sum_{(x',y')\in\Omega}\theta_0-\theta_{c_i}(x',y')\right], \qquad (10)$$

where θ_0 is the predicted orientation, the variables (x',y') run over a local neighborhood Ω whose size is N.

From the complex filter responses on eight channels, the magnitude m and the orientation θ are trivial to compute, and the details are skipped (see, *e.g.*, (Haglund and Fleet, 1994)).

4 RESULTS

In this section, we present and evaluate the results of our mechanism on an artificial (section 4.1) and a real scene (section 4.2).

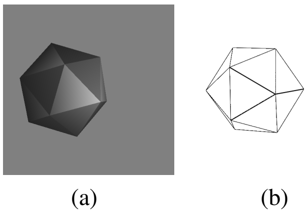

(a) (b)

Figure 6: **(a)** Artificial scene generated using OpenGL. **(b)** Wireframe drawing mode in OpenGL provides ground truth for evaluating the feedback.

4.1 Artificial Scene

The artificial data that we used is icosahedron (*i.e.*, a polyhedron having 20 faces) shown in figure 6(a). The icosahedron is generated using OpenGL which allows us to exercise a certain RBM and make use of the ground truth information to evaluate the performance. The ground truth is computed using the *wireframe* drawing mode in OpenGL (shown in figure 6(b)). We define a feedback true-positive if the image point is close to an edge of the wireframe (namely, the distance is less than three pixels); a feedback is false-positive if it is not a true-positive.

Figure 7 shows the results on the artificial scene. We see in figure 7(e) that many of the 2D primitives are not extracted due to the low contrast. However, knowing the RBM allows the missing edges in figure 7(e) to be extracted with feedback from RBM as shown in figure 7(f).

In figure 8, the improvement of the feedback mechanism is evaluated using the ground truth values. The ROC (Receiver Operating Characteristics) curve in figure 8(a) shows that the proposed feedback mechanism produces a better true to false positive ratio than without the feedback mechanism. In figures 8(b) and (c), the true and false positives on the original image, respectively without and with the feedback mechanism, are displayed (the false-positives are due to shading as shown in figure 2). We see that the feedback mechanism increases the amount of the true positives while decreasing the false positives. Note that the false positives are mostly due to shadows in homogeneous areas of the icosahedron, which sometimes produces edge descriptors which are instable over time. The amount of the true and false positives for different energy (*i.e.*, magnitude) thresholds are displayed in figures 8(d) and (e). A threshold value n means that only edge descriptors whose energy is below n are considered for the evaluation. For example, a threshold of $n = 1.0$ means that all descriptors (edge and non-edge) are included. We see that at all energy thresholds, the feedback mechanism produces a higher true-to-false positive ratio.

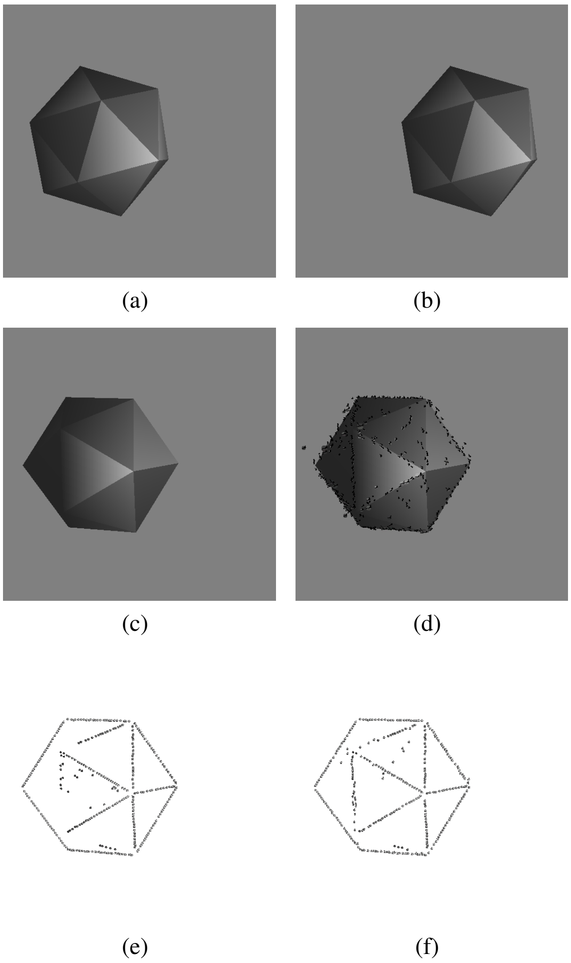
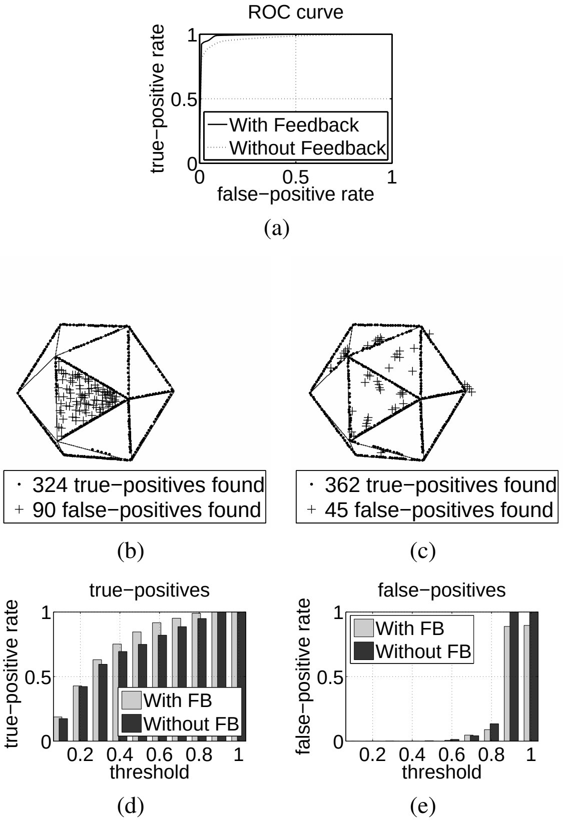

(a) (b) (c) (d) (e) (f)

Figure 7: **(a)-(b)** Left and right frames at time t. **(c)** Left frame at time $t+1$. **(d)** Image projection of the predicted 3D primitives in frame $t+1$. **(e)** 2D primitives extracted in frame $t+1$ without feedback. **(f)** 2D primitives extracted in frame $t+1$ with feedback.

4.2 Real Scene

The real scene involves a robot arm and an object grasped by the robot arm (figure 9). The robot arm executes a known RBM, and our system uses the RBM to improve the feature extraction.

Figure 10(a) shows the extracted primitives without feedback. We see that some of the edges are not extracted due to low contrast. However, the knowledge of RBM can feed back and improve the extraction of the edges (figure 10(b)). Figures 10(c) and (d) show that the extraction of the magnitude is improved with the feedback.



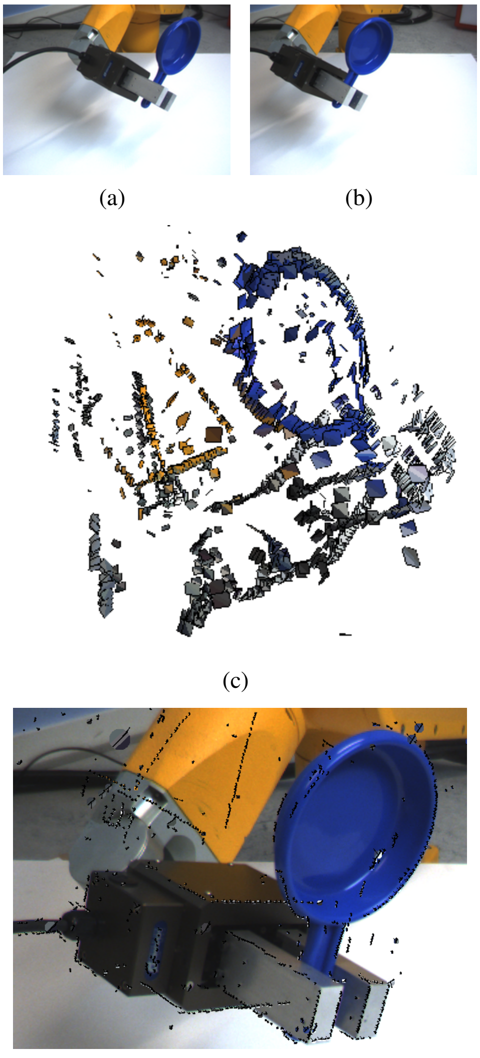
(a) (b) (c) (d) (e)

Figure 8: **(a)** ROC curve for artificial scene. **(b)** True and false positives for primitives whose magnitude is above a magnitude threshold of 0.4 *without* feedback. **(c)** True and false positives for primitives whose magnitude is above a magnitude threshold of 0.4 *with* feedback. **(d)** True positives for primitives with and without feedback for different magnitude thresholds. A threshold n means that only descriptors whose magnitude is below n are considered. **(e)** True positives for primitives with and without feedback for different magnitude thresholds. A threshold n means that only descriptors whose magnitude is below n are considered.

5 CONCLUSION

This paper has proposed a novel feedback mechanism to improve the extraction of low contrast edges. Specific for this mechanism is that information is transformed to a symbolic level on which symbolic reasoning leads to predictions that then become fed back to the signal level. For this, the prediction that has been generated on a symbolic level needs to be inverted to become comparable at the signal level.

In the current paper, symbolic reasoning is restricted to the change of a symbolic descriptor under a rigid body motion. However, we claim that the introduced mechanism is also applicable to other forms



(a)

(b)

(c)

(d)

Figure 9: **(a)-(b)** Left and right frames at time t. **(c)** 3D primitives at time t (extracted from (a) and (b)). **(d)** The projection of the predicted 3D primitives in (c) shown over the image taken at frame $t+1$.

of symbolic reasoning, for example by using stored object knowledge to predict edges at weak structures after an object hypothesis has been aligned with the current scene (as for example in the Dalmatian dog illusion (Gregory, 1970)). These issues are being addressed in our ongoing research.

ACKNOWLEDGEMENTS

We would like to thank Nicolas Pugeault and Emre Başeski for fruitful discussions. This work is supported by the European Drivsco project (IST-FP6-FET-016276-2).

REFERENCES

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

Angelucci, A., Levitt, J. B., Walton, E. J. S., Hupe, J.-M., Bullier, J., and Lund, J. S. (2002). Circuits for Local and Global Signal Integration in Primary Visual Cortex. *J. Neurosci.*, 22(19):8633–8646.

Bayerl, P. and Neumann, H. (2007). Disambiguating visual motion by form–motion interaction — a computational model. *International Journal of Computer Vision*, 72(1):27–45.

Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews*, 36:96–107(12).

Faugeras, O. (1993). *Three–Dimensional Computer Vision*. MIT Press.

Galuske, R. A. W., Schmidt, K. E., Goebel, R., Lomber, S. G., and Payne, B. R. (2002). The role of feedback in shaping neural representations in cat visual cortex. *Proceedings of the National Academy of Science*, 99:17083–17088.

Gonzales, R. and Woods, R. (1992). *Digital Image Processing*. Addison-Wesley Publishing Company.

Gregory, R. L. (1970). *The intelligent eye*. McGraw-Hill Book Company, New York.

Haglund, L. and Fleet, D. J. (1994). Stable estimation of image orientation. In *ICIP (3)*, pages 68–72.

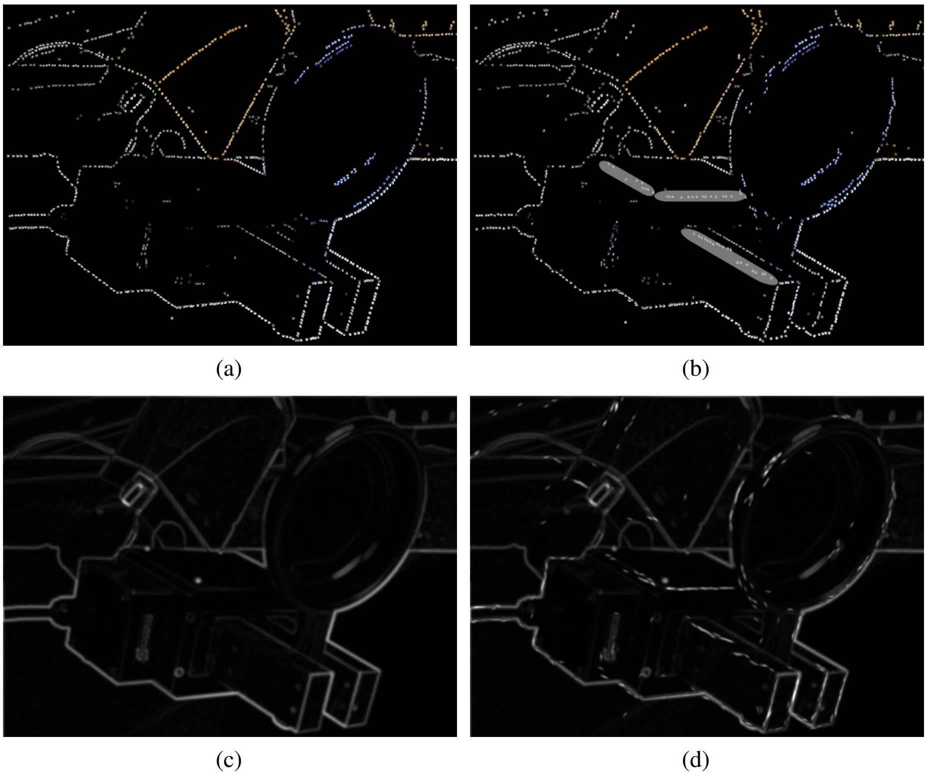
Hubel, D. and Wiesel, T. (1969). Anatomical demonstration of columns in the monkey striate cortex. *Nature*, 221:747–750.

Krüger, N. (2005). Three dilemmas of signal- and symbol-based representations in computer vision. *Workshop on Brain, Vision and Intelligence, BVAI, Naples, Italy*.

Krüger, N., Lappe, M., and Wörgötter, F. (2004). Biologically motivated multi-modal processing of visual primitives. *The Interdisciplinary Journal of Artificial Intelligence and the Simulation of Behaviour*, 1(5).

Lee, T. S. (1996). Image representation using 2d gabor wavelets. *IEEE Trans. Pattern Anal. Mach. Intell.*, 18(10):959–971.

Nixon, M. S. and Aguado, A. S. (2002). *Feature extraction & image processing*. Butterworth Heinmann/Newnes.



(a)

(b)

(c)

(d)

Figure 10: **(a)** The primitives extracted at frame $t+1$ *without* feedback. **(b)** The primitives extracted at frame at $t+1$ *with* feedback. The gray area denotes the extracted descriptors which are lost without feedback mechanism. **(c)** The magnitude image of frame $t+1$ *without* feedback. **(d)** The magnitude image of the updated frame at $t+1$ *with* feedback.