Brain Inspired Cognitive Systems
August 29 – September 1, 2004
University of Stirling, Scotland, UK

NEURAL MULTILAYER STRUCTURE FOR MOTION PATTERN SEGMENTANTION

Eduardo Ros
Departamento de Arquitectura y Tecnología de Computadores, E.T.S.I. Informática, Universidad de Granada, Periodista Daniel Saucedo Aranda s/n, 18071 Granada, Spain
eros@atc.ugr.es

Javier Díaz and Sonia Mota
Departamento de Arquitectura y Tecnología de Computadores, E.T.S.I. Informática, Universidad de Granada, Periodista Daniel Saucedo Aranda s/n, 18071 Granada, Spain
{ jdiaz, smota}@atc.ugr.es

ABSTRACT

Bio-inspired energy models compute motion following the guidelines suggested by the neurophysiological studies of V1 and MT areas of monkeys and human behaviour that use neural populations to extract the local motion structure through local competition of MT like cells. In this paper we present a neural structure that works as dynamic filter on the top of this MT layer for image segmentation and can take advantage of the neural population coding in the cortical processing areas. The test bed application addressed in this work is an automatic watch up system for overtaking situations seen from the rear-view mirror. The ego-motion of the host car induces a global motion pattern whereas an overtaking vehicle produces a motion pattern highly contrasted with this global ego-motion field. We described how a simple neural processing scheme can take full advantage of this motion structure for segmenting overtaking cars in this scenario.

INTRODUCTION

Motion processing is a relevant task for the survival challenge of most of living beings therefore their visual systems have specific areas for motion processing [1]. Primary visual areas are modeled using spatio-temporal receptive filters to compute motion [2, 3, 4] as suggested by neuro-physiological data [5].

We use an energy model based on the work of Simoncelli and Heeger (S&H) that has strong neurophysiological bases. They modeled how the cortical areas (V1 and MT cells) can extract the motion structure through neural local computation and competition. With this model they obtained results that agreed with neuro-physiological data [4, 6]. The output layer uses neural velocity population coding, which is inefficient compared with more mathematical based algorithms but represents and advantage if the post-processing is done through neural computation as presented in this paper.

The MT cells are highly sensitive to a very specific movement direction and speed. This characteristic is based on a high interconnectivity among the cortical layers. Hence it produces smooth and homogeneous motion patterns. We propose a post-processing structure that takes advantage of these properties. Motion estimation based on local operators is normally very noisy and requires further post-processing before addressing any segmentation. In this paper we describe how a simple connectivity pattern facilitates the neural computation of noisy motion information. This connection pattern makes individual cells behave as dynamic filters that are sensitive to more reliable movement features than simple spatio-temporal correlations.

This post-processing layer is composed by cells that collect the output activity from MT cells sensitive to similar motion primitives. We also describe how this can enhance the capability of segmenting rigid body motion by connecting MT cells of local neighbourhoods throughout the visual field.

The application of this neural processing strategy in real world problems is also illustrated. In particular, promising results have been obtained for an overtaking car segmentation task. This problem is currently being addressed by many application driven research groups [7]. Besides, in this scenario the motion processing plays an important role, since an overtaking car exhibits a forward motion pattern clearly contrasted against the global backward motion pattern observed in the rear-view mirror due to the ego-motion of the host car.

VELOCITY ESTIMATION USING A NEURONAL COMPUTATION SCHEME

The Simoncelli & Heeger model [4, 6], consists of two primary stages corresponding to cortical areas V1 and MT. The computation in these layers is highly parallel and regular.

A linear model is used for V1 simple cells that exhibit specific selectivity for stimulus orientation and spatial frequency.

A basic set of tuned V1 neurons covers a wide range of spatio-temporal frequencies with low overlapping. Each V1 neuron squares and normalizes its inputs. The next neural layer models V1 complex cells. They receive afferents from V1 simple cells distributed over a local spatial region, sharing the same space-time orientation and phase. This forms the V1 complex cells receptive field. In this way V1 complex cells compute a weighted sum of these inputs. In other models energy neurons are modeled using quadrature Gabor filters [8, 9]. Our approach uses receptive filters based on third Gaussian derivatives and spatial pooling as suggested by Simoncelli et al [4], (see Fig. 1).

MT cells are modeled combining the outputs of a set of direction-selective V1 complex cells, whose preferred space-time orientations are consistent with the MT cells characteristic velocity. The mechanism for velocity selectivity can be described in the spatio-temporal domain easily. The power-spectrum of a translational pattern lies on a plane, and the tilt of the plane depends on the velocity. In this way a MT cell detects the tilted plane with maximum response [10]. The MT cell uses the weighted V1 inputs and interpolates the different spatio-temporal planes to tune their preferred plane. Different combinations of V1 cells can be used to form the MT receptive field [3, 11]. In our approach, a mechanism based on vector projection is used to obtain the interpolation weights.

Finally, a Winner-Takes-All configuration among the MT population selects only the MT cells with higher input, i.e., the ones that best match the local motion pattern as shown in Fig. 2.

The implementation of S&H model, showed in Fig. 3, can be summarized in 4 steps:

1. Compute local contrast stimulus.
2. Model V1 simple and complex neurons, using spatio-temporal third Gaussian derivatives and spatial pooling.
3. Model MT neurons summing the weighted responses of V1 cells which lie on its characteristic plane.
4. Compute a Winner-Takes-All to select a single cell for each pixel in the visual field

Our approach uses a basic spatio-temporal set of 40 filters (8 spatial and 5 temporal orientations) and a single spatial scale. Gaussian derivatives are preferred over Gabor filters because they are steerable filters [12]. Only 10 convolutions are need to calculate the 40 spatio-temporal orientations instead of the 40 convolutions needed for Gabor filters. With this limited set of V1 orientations we tune a set of 121 MT cells with different preferred velocities. The final system is the pyramidal structure shown in Fig. 4, were a population of neurons tune different motion patterns.

Other important topic is the contrast dependency of energy models [2]. The neuron model used has the capability of auto-normalization [8] and with the competition layer scheme this problem is minimized.

One limitation for this neural structure is that it can not detect second order motion. Some modifications could be added to detect it [13], but this is out of the scope of this contribution. Furthermore the addressed real-world application requires mainly accurate translational motion processing thus second order motion detection is not required.

COLLECTOR LAYER

Now we describe a simple neural structure that can take advantage of the population coding at the MT layer for a specific application such as the segmentation of overtaking cars.

The MT layer is connected to a new neural layer that we call Collector Layer (CL). The cells at this stage receive excitatory convergent many-to-one connections from the MT layer. The CL cells work as filters which efficiently segment rigid-bodies. Each CL cell is sensitive to a set of velocities $\mathbf{V} \pm \Delta\mathbf{V}$ from MT outputs, where $\Delta\mathbf{V}$ represents slight variations in module and angle from preferred values, i.e. each CL neuron integrates the activity of those MT cells in a spatial neighbourhood that tune the characteristic velocity of this CL neuron

The main task of the proposed scheme is the improvement of rigid-body motion detection. There are different CL neurons in the same area of the collector layer and each one is tuned to a different set of velocities. The CL is configured as a self-competitive layer, i.e. the collector neuron that receives the maximum contribution in its spatial influence area inhibits the others and dominates in this area (Winner-Takes-All). This helps to detect rigid body motion. If we neglect possible rotations that are only of marginal importance for overtaking scenes, all points of a rigid body share the same speed and motion direction. Isolated points belonging to a rigid body that move with other velocities are considered noise.

As the application addressed is focussed on discriminating between leftward (ego-motion) and rightward (overtaking vehicle) moving features, only the cortical S&H neurons that match these directions are connected to the Collector Layer for this task.

The configuration of the collector layer neurons can embody another important aspect for the segmentation task: perspective deformations of motion patterns.

Due to this effect, an overtaking vehicle although moving at a constant speed, seems to accelerate as it approaches, i.e. it is expected to move slowly when it is in the very left side of the image (far away) and its speed increases when it moves rightwards to a closer position. To reduce this effect, the distribution of the specialised collector neurons is non-uniform. The ratio of cells tuned to high speeds is lower in the left side of the visual field than in the right side. The opposite is done with cells more sensitive to slower speeds. This facilitates the detection of slow movements in the left side of the visual field and rapid movements on the right side. This reduces the effect of the perspective deformation.

The same perspective problem damages the perception of moving solid objects, because the overtaking vehicle rear and front ends are moving at different speeds. That is critical for very close vehicles. The sensitivity of each CL cell to a set of characteristic speeds instead of a single one, corrects this perception problem.

On the other hand, the winner neurons in a local influence area at CL compete locally with other winner neurons from other areas in the neighbourhood. This interaction facilitates the domination of large features and inhibits those winner neurons whose motion direction is different with respect to the majority

of the surrounding winner cells. In this way, the output response of this filtering neural layer (CL) will be non-zero if there are winner collector neurons non-inhibited by others winner cells (Fig. 5). In Fig. 5 C2 is inhibited by C3 because they have opposite direction selectivity. But C1 and C2 receive cross excitation because their selectivity characteristics are coincident. This enhances coherent patterns moving and reduces fuzzy estimations.

Other property of CL neurons is a time constant that takes into account how the stimulus drives the onset and offset of the elements of this layer. If we choose this time constant to be long, that means that more integration time from a lasting motion pattern is needed to activate a neuron and make it dominate against previously detected patterns. This also improves the stability of response for translational motion pattern in noisy environments and reduces the velocity deformation due to the perspective.

EVALUATION RESULTS

The described neural system has been applied to four real overtaking sequences. The results are summarized in Fig. 6. The proposed neural processing scheme segments efficiently rigid objects that are moving in opposite horizontal directions.

Fig. 6.a and Fig. 6.b show an overtaking sequence with a dark car in a sunny day recorded with a conventional CCD camera. The other sequences were taken with a HDR (high-dynamic-range) camera [14]. These sequences are: an overtaking truck (Fig. 6.c), single overtaking car in a foggy and rainy day (Fig. 6.d) and a sequence of multiple overtaking cars (Fig.6.e) with some mist.

Fig. 6 is distributed in columns. The left column shows an original image of the overtaking sequences. The middle column shows the S&H extracted optical flow. The arrows show the motion direction (arrow sizes do not contain additional information due to the large range of velocities present in the sequences) and the grey scale indicates the speed (lighter colour indicates faster motions). The right column shows the CL outputs. The segmented overtaking car is represented with a dark colour (rightward motion) and the background, moving to the opposite direction, uses bright colour.

The receptive fields of the collector layer receive connections of MT neurons tuned to a cone of velocity directions mainly focused in horizontal motions. Due to this, as can be seen in the bottom part of the car in Fig. 6.b, the optical flow out of this cone is neglected by the collector layer.

Some weather conditions (fog and rain) reduce the contrast of the sequences while the lights of the cars would easily saturate CCD sensors; therefore open-air applications usually require HDR cameras. In spite of the use of these cameras, the extracted optical flow is worse in adverse weather conditions than the one of a sunny day sequence, reducing the confidence of motion discrimination. Other effects that lead to worse car segmentation are reflections of light on the road and noisy artefacts produced by the rain.

The high dynamic range camera generates 32 bits precision while our model works with 8 bit-depth. This precision restriction induces other artefacts that lead to wrong velocity estimations that are very significant in low contrast sequences such as shown in Fig. 6.d. Nevertheless, the proposed neural computing scheme efficiently deals with all these artefacts in overtaking scenarios.

CONCLUSIONS

This paper describes a bio-inspired system used to segment objects through motion energy extraction. A post-processing layer (the collector layer) filters the motion information of MT layer. The CL connection topology embodies aspects that facilitate the segmentation of moving rigid-bodies and reduce the effect of the perspective deformation of the visual field due to the rear view mirror.

The proposed neural system is highly parallel. It is a self-competitive neural computation scheme for feature selection. This enhances the capability of segmenting rigid bodies in noisy environments as seen in Fig. 6.

ACKNOWLEDGMENTS

This work has been supported by the V EU research framework funds through the European Projects ECOVISION (IST-2001-32114) [15] and CORTIVIS (QLK6-CT-2001-00279). We would like also to acknowledge Hella [16] for providing us with the overtaking sequences.

REFERENCES

1. K. Nakayama: Biological image motion processing: a review. *Vision Research*, 25, pp. 625-660, 1985.

2. E. H. Adelson and J. R. Bergen: The extraction of spatio-temporal energy in human and machine vision. *In Proc. of IEEE Workshop on Motion*, pp. 151–156, 1986.

3. D. J. Heeger: Model for the extraction of image of Image flow. *Journal of the Optical Society of America* A4, pp. 1455–1471, 1987.

4. E. P. Simoncelli and D. J. Heeger: A model of Neuronal Responses in Visual Area MT. *Vision Research*, 38(5), pp. 743-761, 1998.

5. D. H. Hubel and T. N. Wiesel: Receptive fields, binocular interactions and functional architecture in the Cat's Visual Cortex. *J. Physiology*, 160, pp. 106-154, 1962.

6. E. P. Simoncelli: Distributed Analysis and Representation of Visual Motion. *PhD thesis, Massachusetts Institute of Technology, Dept of E. Eng. Comp. Sci., Cambridge,MA.*, 1993.

7. U. Franke, D. Gavrila, A. Gern, S. Görzig, R. Janssen, F. Paetzold and C. Wöhler: From door to door- Principles and Application on Computer Vision for driver assistant systems, in *Intelligent Vehicle Technologies: Theory and Applications*, Arnold 2000.

8. D. J. Heeger: Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9, pp. 181–198, 1992.

9. D. J. Fleet, H. Wagner and D. J. Heeger: Neural encoding of binocular disparity: Energy models, position shifts and phase shifts**.** *Vis. Research,* 36(12), pp. 1839-1857, 1996.

10. A. B. Watson and A. J. Ahumada: A look at motion in the frequency domain, in *Tsosos, J.K., editor, Motion: Perception and representation*, pp. 1–10. NewYork, 1983.

11. N. M. Grzywacz and A. L. Yuille: A model for the estimate of local image velocity by cells in the visual cortex. *Proceeding of the Royal Society of London* A 239, pp. 129–161, 1990.

12. W. T. Freeman and E. H. Adelson: The design and use of steerable filters. *IEEE Pattern Analysis and Machine Intelligence*, 13, pp. 891–906, 1991.

13. G. Sperling, C. Chubb, J. A. Solomon and Z-L Lu: Fullwave and halfwave processes in second order motion and texture, in Higher-order processing in the visual system. Chichester, U.K., Wiley (Ciba Foundation Symposium, 184), pp. 287-303, 1994.

14. IMS-CHIPS: http://www.ims-chips.de/home.php3?id=e0841.

15. ECOVISION, http://www.pspc.dibe.unige.it/~ ecovision/

16. Dept. of predevelopment EE-11, Hella KG Hueck & Co., Germany, www.hella.de

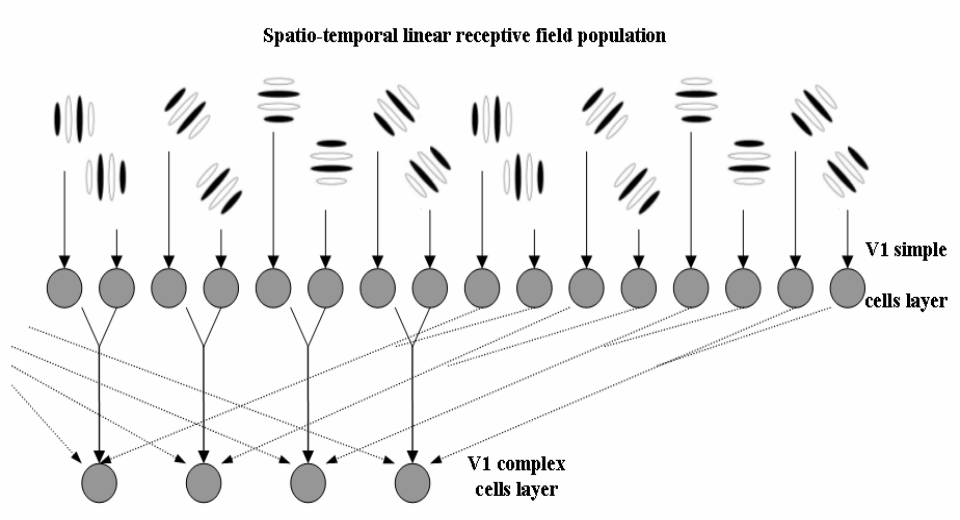

Fig. 1. V1 simple to complex cells interconnections. V1 complex cell are modeled using a Gaussian pooling operation.

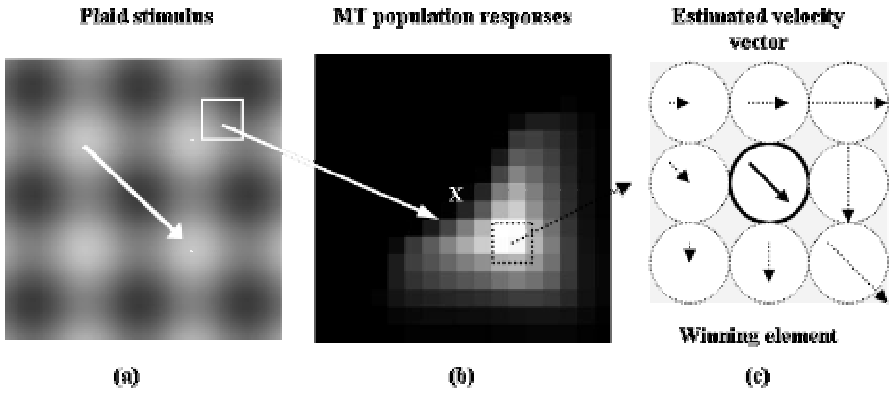

Fig. 2. Example of population coding. (a) The result of a plaid stimulus composed by a sinusoidal grating moving rightwards and another moving downwards is a moving pattern toward the right-bottom corner. (b) Grey levels codify the responses of a set of MT neurons. The relative position of the winner element with respect to the center of the population encodes the velocity module and direction. Maximum responses are given at the best tuned MT neuron for that stimulus, but MT cells tuned to near velocities are not zero. (c) Finally, the winner element characteristics encode the estimated velocity.

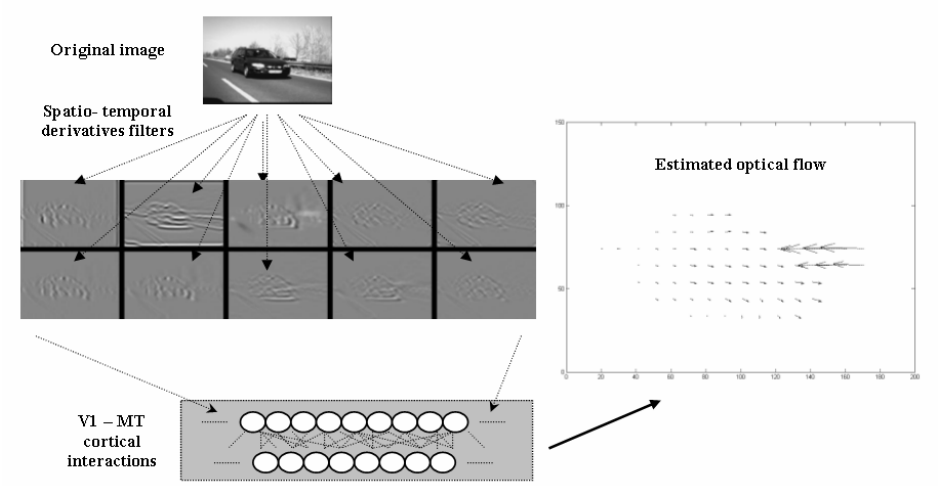

Fig. 3. S&H Model. An overtaking car sequence is used to evaluate the model. After convolution operations with Gaussian derivative filters, the pre-filtered images are combined to get V1 spatio-temporal orientation cell responses. These are combined to obtain the MT cell output. Finally, after a winner-takes-all competition process only one neuron per pixel remains active whose inherent characteristics encode the estimated velocity vector.

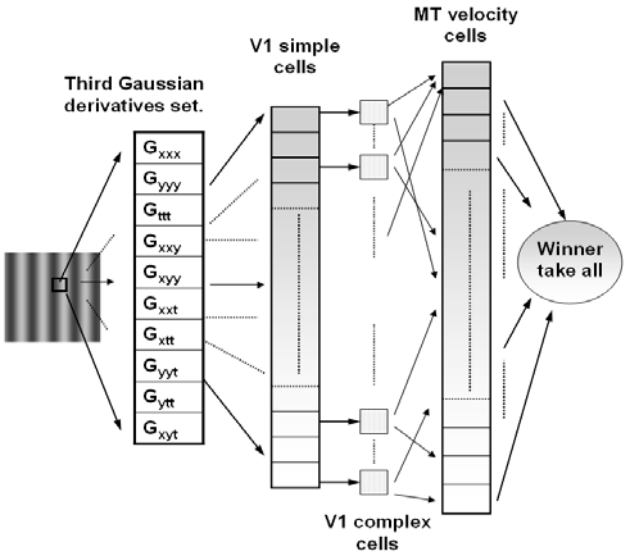

Fig. 4. Pyramidal neuronal structure.

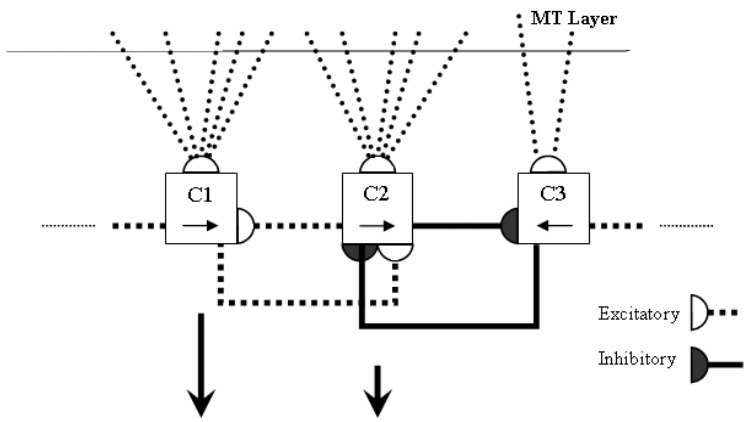

Fig. 5. The figure shows the synaptic connections between three winner collector neurons that integrate the activity of MT cells of similar characteristics over a spatial neighbourhood. Two neurons detect rightward motion direction (→) and the other detects leftward motion detection (←).

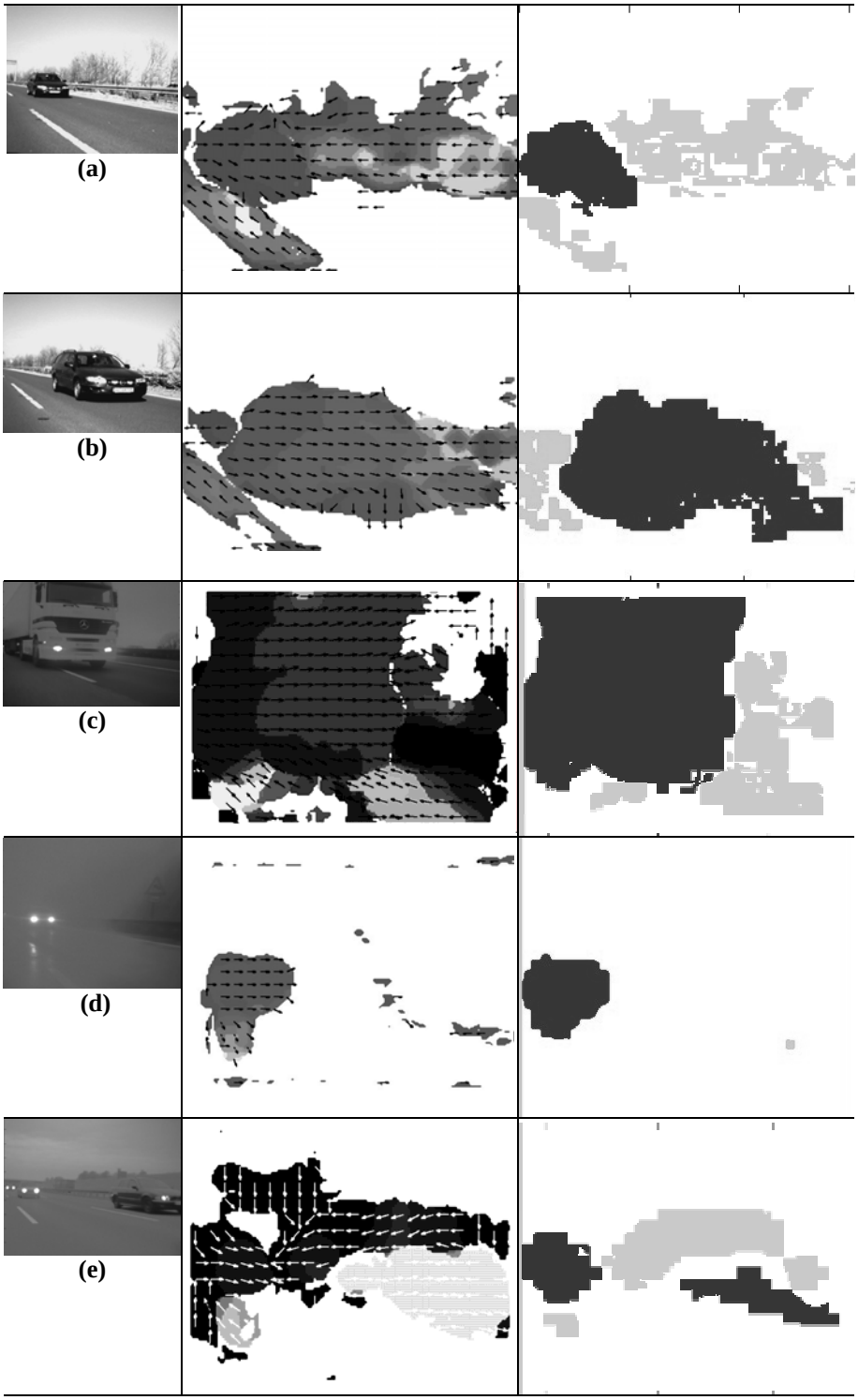

Fig. 6. Overtaking car sequence in a sunny day (a,b); in a cloudy day with mist (c,e); in a foggy and rainy day (d).