Fatigue Features Based on Eye Tracking for Driver Inattention System

Karel Horak

Abstract—**This paper deals with segmentation methods and fatigue features determination for a camera-based visual systems monitoring driver vigilance. Generally visual monitoring systems have to analyse a set of computed fatigue features and recognize driver inattention or sleepiness. The paper is focused mostly on the segmentation methods used for reliable eyes tracking because of eyes features are certainly the most significant features for determining of a driver fatigue. Fundamentals segmentation methods as a simple colour segmentation and Hough transform are introduced in the paper. After that a more complex Haar-like features approach and symmetries detection approach are shortly introduced. Finally, several of the leading fatigue features are listed and described. All the presented segmentation methods were tested on both laboratory and real images.**

***Keywords*—colour segmentation, eye tracking, fatigue features, Haar-like features, face symmetries.**

I. Introduction

ACCORDING to traffic statistics, thousands fatalities happen in motor vehicles every year due to driver inattention and lack of sleep [1]. It follows that real-time determination of driver inattention is a crucial step on the path to increasing a traffic safety. The camera-based visual systems are designed most often alongside of other methods as an electroencephalographic scanning, evaluation of steering wheel movements etc. Each visual system employs more or less different computer vision approach, different hardware platform and different image processing methods for object segmentation and classification. This article focuses on fundamental image processing methods for an estimation of the driver fatigue. The paper mainly deals with colour segmentation approach, the Hough transform, the Haar-like features and symmetric features. Furthermore, an image acquisition platform is briefly described in the second chapter and several most significant fatigue features are introduced in the fourth chapter before conclusion.

A. *Traffic Safety*

From a general perspective a traffic safety can be seen as either a common or an individual safety. First mentioned common safety includes particularly a traffic management as traffic lights control, congestions management etc. More interesting for this paper is the individual traffic safety, which covers only systems of a single vehicle. The individual traffic safety can be further considered as passive or active. Although the terms passive and active safety are trivial in general they are very important in the world of automotive safety.

The term passive safety refers to all components inside the vehicle that help to protect passengers during a crash. Examples of such components are airbags, seatbelts, a physical structure of the vehicle etc. On the contrary, the term active safety is commonly used to refer to one or more technology assisting in the prevention of a crash, in other words active systems works prior to an accident.

Each active technology as defined in the previous paragraph can evaluates either external or internal event and so it is referred as an external or internal technology respectively. Most recent systems as lane assist, traffic sign recognition and pedestrian detection are examples of such external active technology. Estimation of the driver fatigue (also inattention, vigilance or sleepiness) by means of the camera-based visual system is the clear example of an internal active technology.

B. *Proposed Solution*

The basic framework for designing of the camera-based visual system intended for monitoring driver vigilance is introduced in this paper. The proposed approach is based on detecting and tracking the driver's eyes while driving a car. Several different approaches are employed for the image segmentation task due to the comparison of their suitability.

The article is ordered as follows. The following chapter briefly describes image acquisition issues as an image acquisition platform, camera parameters and human eye parameters. Third chapter is focused on segmentation techniques, which process image in such a way that vector of fatigue features can be subsequently easily constructed. Selected fatigue features and construction of the feature vector is basically introduced in the fourth chapter and a conclusion is given at the very end of the article.

II. Image Acquisition

A. *Acquisition Platform*

There is a wide range of suitable office-based and industrial cameras as well as appropriate lens in the market. For the purposes of this article is not important what interface the camera is equipped. All testing and verification images were captured by the industrial colour camera ProsilicaGC with the GigE interface, resolution of 1280 by 960 and maximal frame rate of 32 frames per second at full resolution. If necessary, decreasing the image resolution allows proportionally

Manuscript received May 17, 2011. This work was supported by the Ministry of Education of the Czech Republic under the project MSM0021630529 Research Intention Intelligent Systems in Automation and project 1M0567 Centre for Applied Cybernetics. Both are very gratefully acknowledged.

K. Horak is with the Department of Control and Instrumentation, Brno University of Technology, Brno, 61200, Czech Republic (phone: 00420-5-4114-1157; fax: 00420-5-4114-1123; e-mail: horak@feec.vutbr.cz).

increasing the frame rate. A very simple and compact image acquisition platform is strictly required for non-intrusive implementation of the visual system for monitoring driver vigilance inside a car. In some previous works an extra IR illuminator was used to highlight driver pupils especially at night [2]. Using the ring-shaped IR illuminator yielded in easier and more effective eyes detection. Nevertheless, for the reason mentioned above only the simple camera without any extra illuminator was employed in the subsequent research [3]–[4].

Like the resolution or the frame rate of the used camera, an exposure time has to be handled otherwise under-exposed or over-exposed images are delivered from the camera. To prevent this, the two different algorithms (one linear and one logarithmic) were proposed in previous works [3]–[4]. Note that all the industrial cameras allow automatic exposure, however such mode often results in unsuitable images.

B. Acquisition Timing

Acquisition timing acts a crucial role throughout image processing and driver fatigue monitoring. Experiments showed the normal eye closure doesn't exceed 200 milliseconds. In contrast, eye closure of a tired driver often exceeds 1.5 second.

If we consider above mentioned camera of 32 fps the normal blink is captured by at least five frames as can be seen in the Fig. 1. On the contrary an abnormal blink of the tired driver is captured even by approximately fifty frames. It means that image processing algorithms should have sufficient amount of an input data to detect the sleepiness.

III. Segmentation Methods

There is a very large set of segmentation methods in the field of computer vision [5] but only a few of them are suitable for purposes of the driver vigilance monitoring [6]. Several basic image processing methods often used to detect and to track driver's eyes are introduced in the following chapters.

A. Colour Segmentation

Colour segmentation represents class of methods which analyses colour space of an image. A native format of colour industrial cameras is the Bayer format i.e. colour channels are represented as R, G and B matrices. Such representation is not suitable for colour analysis so that a transformation to the YCbCr colour space is carried out and a sub-space CbCr is then employed in colour analysis.

Let I(x,y) refers to an image data after transformation to YCbCr colour space as described ni the previous paragraph. In the training stage all the image pixels from I(x,y) belonging to a human face are selected and the representative vector $(C_B,C_R,\sigma_B,\sigma_R)$ is created.

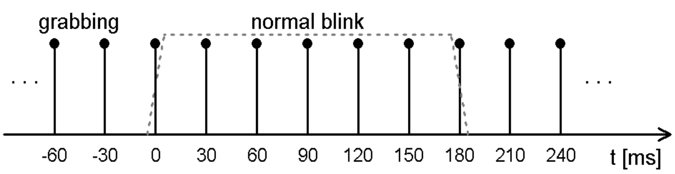

Fig. 1. Normal blink captured by the camera.


Fig. 2. The original input image (top) and the probabilistic map (bottom).

Symbols C_B and C_R stand for central values of blue and red component corresponding to the face colours respectively. Because of a probabilistic approach of segmentation both symbols σ_B and σ_R stands for standard deviations of the two-dimensional normal distribution in the CbCr plane. In the classification stage a membership value p(x,y) is evaluated for each pixel from an input image by means of the relation (1).

$$p(x,y)=\frac{1}{2\pi\sigma_B\sigma_R}\cdot\exp\left[-\frac{(I_B(x,y)-C_B)^2}{2\sigma_B^{\ 2}}-\frac{(I_R(x,y)-C_R)^2}{2\sigma_R^{\ 2}}\right] \tag{1}$$

After that a probability map can be simply constructed as an image matrix of probabilities p(x,y). Both the input image and the probability map are showed in the Fig. 2.

In equations (1) and (2) symbols I_B and I_R denote blue and red channel in the transformed input image respectively. In addition to the face colour detection by means of probability map, the FLD transformation for mouth detection or the so called EyeMap transformation for eyes detection can be executed [7]. The EyeMap colour transformation is given by a simple formula (2) which results to an image as can be seen in the following Fig. 3. The simple eye tracker is then achieved by a combination of the probabilistic map and the EyeMap.

$$m(x,y)=\frac{1}{3}\cdot\left(I_B^{\ 2}(x,y)+neg(I_R^{\ 2}(x,y))+\frac{I_B(x,y)}{I_R(x,y)}\right) \tag{2}$$

Note that function neg(A) stands for an image inversion i.e. a complement of the image function A.


Fig. 3. The probabilistic image of the original input image.

The main advantage of such approach is algorithm simplicity and fast. On the contrary a low robustness and accuracy of implemented algorithm can appear when an ambient illumination strictly differs from the training one.

B. Hough Transform

The Hough transform is usually used to detect geometrical entities as lines, circles, ellipses and other generalized shapes. The Hough transform for detection circle-shaped objects is here employed for localization of driver's eyes.

First step is to compute edges in an input image. This can be done simply by means of convolution input image with e.g. Sobel mask or also by means of Canny edge detector. Because of high computational complexity of the Hough transform it is recommended to reduce an edge image to only face region before transformation. After the mentioned reduction the Hough space (sometimes called accumulator) can be constructed and filled by a voting system. It means that all the pixels corresponding to some circle in the edge image vote for the one point in Hough space. In our case the Hough space is the three-dimensional space of parameters r, φ and R. The first two symbols r and φ are the parameters of circle in the parametric form and the third symbol R denotes radius of the searched circle i.e. radius of driver iris. To detect driver eyes is needed to find the two largest peaks in the Hough space. Coordinate triplet (r,φ and R) of each peak represents one driver eye and determine its coordinates (x,y) in the Cartesian coordinate system of the original input image by the following formula (3).

$$\begin{bmatrix} x \\ y \end{bmatrix} = r \cdot \begin{bmatrix} \cos\varphi \\ \sin\varphi \end{bmatrix} + \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \tag{3}$$

A significant advantage of the Hough transform is its strong insensitivity to noise in the input image because voting style. Second major advantage is its insensitivity to incompleteness of an input data so that even iris of a half-closed eye can be detected and localized. On the contrary, as mentioned above the main drawback is the high computational complexity of the Hough transform algorithm.

C. Haar-like Features

The Haar-like features are based on fundamentally different approach than colour segmentation described above. The theory of the Haar-like features employs Haar basis functions which have been originally used by Papageorgiou for object detection [8]. Within the meaning of a face recognition Viola and Jones introduced a face detection framework that is capable of processing images extremely rapidly while achieving high detection rates [9]. The face detection algorithm only classifies images based on values of simple features. A basic set of the Haar-like features contains four rectangle features as it is shown in the following Fig. 4.

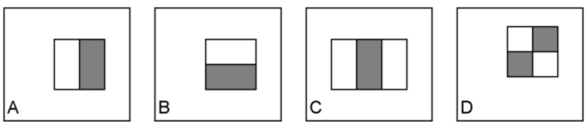

Fig. 4. The basic set of the Haar-like features for rapid object recognition.

Values of the selected Haar-like features A, B, C and D is computed simply as the sum of the pixels within the corresponding bright rectangle subtracted from the sum of pixels within the corresponding dark rectangle. A so-called two-rectangle features in (A) and (B) detect the vertically and horizontally adjacent regions with different brightness values. The three-rectangle feature (C) computes the sum of all pixels within two outside bright rectangles subtracted from the sum in a centre dark rectangle. Finally a four-rectangle feature (D) computes the difference between diagonal pairs of corresponding rectangles.

The feature A, B, C or D (or some combination of them) is computed for each pixel of an input image as a difference between values of dark and bright regions. More specifically, as mentioned above the value of a so-called two-rectangle feature (Fig. 4 A and B) is given by the difference between the sum of pixels within two corresponding rectangular regions. Owing to high computational cost of mechanical calculation (pixel by pixel), an intermediate image representation called the integral image has to be used for speed up an algorithm. The integral image at location (x, y) is given by the sum of the all image pixels above and to the left of coordinates (x, y). Accurately the integral image $I_{int}(x,y)$ of an arbitrary greyscale input image I(x,y) is given by the formula (4).

$$I_{\text{int}}(x,y) = \sum_{i \le x, j \le y} I(x,y) \tag{4}$$

The significant advantage of the integral image $I_{int}(x,y)$ is computation in one pass over the original input image I(x,y) as shows the following two relations (5):

$$\begin{aligned} s(x,y) &= (x, y - \;) + \; (x,y) \\ I_{\text{int}}(x,y) &= {}_{\text{int}}(x - \;, y) + \; (x,y) \end{aligned} \tag{5}$$

where s(x,y) denotes an intermediate sum of all pixels in column x with a vertical coordinate less than y.

As mentioned above the integral image $I_{int}(x,y)$ refers to sum of all pixels above and to the left of coordinates x and y respectively as depicted in the Fig. 5. The Haar-like features can be constructed by means of the integral image very rapidly. For example the sum of all pixels within rectangle D on the right picture in the Fig. 5 can be easily determined as a sum or a difference of four values. The sum within the rectangle D is equal to 4-2-3+1 where e.g. number 4 refers to the value of integral image at point marked with the number 4 (see the picture on the right in the Fig. 5). Using integral image the Haar-like features can be computed at any scale and any location very rapidly and in constant time.

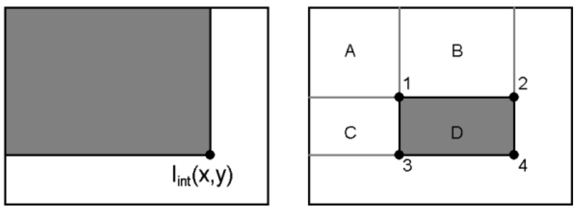

Fig. 5. The integral image and a the sum of pixels inside the rectangle D.

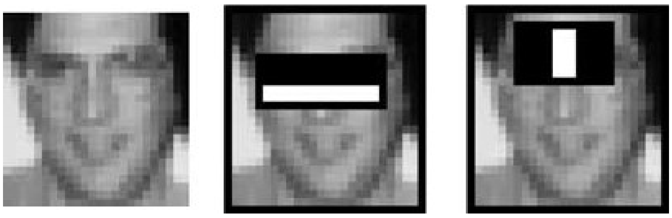

Fig. 6. The positions of the (B) and (C) Haar-like feature localized by classifier based on adaptive boosting technique [9]

The basic set of the Haar-like features described above can be gracefully used for face detection and for face features detection as can be seen in the Fig. 6. When the face or the face feature detected, the face region is divided into two regions of eyes and into a residual region. In most applications a contrast correction has to be applied on the detected eyes regions. After the two eyes regions are obtained, the circular type of the Hough transform can be applied on them for detection of pupils. The circle type of the Hough transform is a variation of the classical Hough transform which was originally intended for an identification of straight lines.

In addition to the basic set of Haar-like features the extended set of the Haar-like features are described in [10]. These new set of rotated features improves face detection and results in approximately 10 % lower false alarm rate.

The face detection procedure just described is suitable for decreasing an area of interest needed for further eye tracking algorithm. Eyes tracker then achieves better results due to the reduced area. Such approach can suppress the main drawback of the Hough transform as mentioned in previous chapter i.e. computational complexity because of only small regions of eyes is further processed. Note that detection the skin of a human face is in general a hard task due to the wide variance of ambient illumination in real images.

D. Symmetric Features

Symmetric features approach has been used for numerous applications including face detection and further facial image analysis [11]. It is a clear idea that a human face is vertically symmetric. It follows the driver face or his eyes can be successfully located by means of a symmetric feature computed on a set of interest points. These can be easily detected using e.g. well-known Moravec or more complex Harris operator. Each image of a human face contains a large number of interest points but only some of them contribute to a value of symmetric feature.

In the first place a set of feature points p_i has to be constructed using an arbitrary rotationally invariant method such as SIFT [12]. A requirement of a rotationally invariance is very important due to a further computation of rotated features for the each interest point whilst scale-invariant parameter is not necessary. Let the p_i be a point vector of the i^{th} interest point defined as (6).

$$\overline{p}_i = (x_i, y_i, \phi_i, s_i) \tag{6}$$

where the pair x_i and y_i denotes location of the interest point, symbol ϕ_i denotes its orientation and symbol s_i its scale.

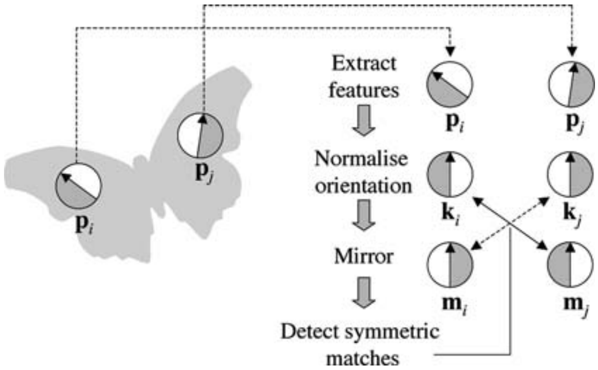

Fig. 7. Illustrating of the descriptor pair (k_i,m_i) and pairs matching [13]

Next a feature descriptor labelled k_i is computed for each feature point p_i as defined above. The feature descriptor k_i encodes the local appearance of the i^{th} feature point after its orientation and scale have been normalised. After that a mirrored feature descriptor labelled m_i is generated for each corresponding feature descriptor k_i. The mirrored feature descriptor m_i is simply the same as basic feature descriptor of mirrored image region. The basic feature descriptor k_i and the mirrored feature descriptor m_i together create a descriptor pair.

As it is shown in the Fig. 7 each pair of a feature descriptor k_i of interest point p_i and a feature descriptor m_j of interest point p_j is matched and a similarity measure M_{ij} is generated. It is obvious that matching the pair (k_j,m_i) is redundant because the similarity measure M_{ij} is the same as M_{ji} due to the relationship to the same pair of interest points. The similarity measure M_{ij} is given by relation (7):

$$M_{ij} = \begin{cases} \Phi_{ij} \; S_{ij} \cdot D_{ij} & \Phi_{ij} > 0 \\ 0 & otherwise \end{cases} \tag{7}$$

where Φ_{ij} refers to an angular symmetry weighting, symbol S_{ij} refer to a scale weighting and finally symbol D_{ij} refers to a distance weighting. When similarity measures are known the linear Hough transform can be employed to determine a dominant symmetric axis. Only similarity measures M_{ij} with value greater than a-priori given threshold are processed in Hough transform voting style. It means that each pair (p_i,p_j) with the strong symmetry measure M_{ij} votes a couple (r_{ij},θ_{ij}) in a Hough accumulator. When finished the Hough accumulator is blurred with a Gaussian filter and a peak corresponding to a global maximum is found so that the specific couple (r_{ij},θ_{ij}) is extracted. This two coordinates clearly determine the dominant symmetric axis in an input image as it is shown in the next Fig. 8.

Fig. 8. The original input image, symmetric feature descriptors and the dominant axis with associated interest points [13].

When determined a dominant symmetry axis the interesting points corresponding to eyes regions have to be easily located and subsequently tracked over the image sequence. A computational complexity of symmetric features approach is in particular given by a computational complexity of interest points detector and features descriptor algorithm.

IV. Fatigue Features

The face detection algorithm and eyes tracking mechanism is only the first stage for driver vigilance determination. A degree of the driver inattention or fatigue is computed from a set of so-called fatigue features. As a fatigue feature can be considered each quantity which follows real driver inattention. Nevertheless only a finite set of significant fatigue features is employed in real applications. The most often parameters used are percent eye closure (so-called PERCLOS), eye closure duration, blink frequency, fixed gaze, head nodding frequency and face position [14]. Only the first four parameters deals with driver eyes and the last two deals with head behaviour.

The PERCLOS parameter refers to a speed of the driver blink. Consider an eye is open if eye-closure is above 80 % of full opened eye and on the contrary the same eye is closed if eye-closure is less than 20 % of full opened eye. See the next Fig. 9 for time periods labelling. The solid black curve illustrates an eye blink and grey dashed lines identify four period t_1, t_2, t_3 and t_4.

The PERCLOS parameter is verbally defined as a ratio of time period when the eye is closed and time period when the same eye is open in sense of definitions above. Exactly is PERCLOS given by the following relation (8).

$$PERCLOS = \frac{t_{eye\_closed}}{t_{eye\_open}} = \frac{t_3 - t_2}{t_4 - t_1} \cdot 100 \quad [\%] \tag{8}$$

It is important that shape of blink curve as depicted in the Fig. 9 does not change when increasing driver fatigue. On the contrary blink duration is gradually decreasing. It follows that blink of a tired driver is much slower than blink of an attentive driver but shapes of eye-closing are almost identical for both cases. Other remaining parameters are treated in a similar way as PERCLOS and all these values are subsequently used as an input for classifier. It means at the very end all the parameters corresponding to fatigue features are combined using a fuzzy classifier or some other technique to specify a degree of driver inattention. Fusion of multiple visual parameters yields higher robustness and less misclassification than by only using one parameter.

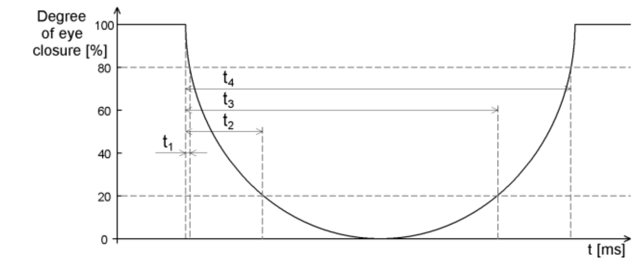

Fig. 9. Illustrating the timing curve of eye blink for PERCLOS parameter

V. Conclusion

The presented paper introduced a formal framework for the camera-based visual systems intended for monitoring of the driver inattention. The main attention was primarily paid to principally different techniques for an image segmentation used for eyes tracking. In this section the colour segmentation approach and the Hough transform have been described as essential segmentation methods as well as more complex techniques the Haar-like features and symmetric features. The main advantage of presented approach is reducing complexity and increasing robustness of the task. Suggested approach is not directly dependent on raw pixel value as in [7] or luminance value (in case of RGB space in colour segmentation) but is based on more independent values as C_bC_r components or edge image. Also hardware platform is compact as possible without any extra IR illuminator as for example used in [14].

Future work will be focused on fatigue features extraction in the sense of real-time and on a classifier enables reliable and efficient determination of the driver fatigue.

References

[1] Smith, P., Shah, M., da Vitoria Lobo, N., Determining driver visual attention with one camera, *IEEE Transactions on Intelligent Transportation Systems*, 2003, pp. 205-218.

[2] Horak, K., Richter, M., Kalova, I., Human Eyes Localization for Driver Inattention Monitoring System, *15th International Conference on Soft Computing*, Brno, 2009, pp. 283-288.

[3] Horak, K., Kalova, I., Image Acquisition and Processing for Monitoring Driver Vigilance, *16th International Conference on Soft Computing, Brno*, 2010, pp. 517-523. ISSN 1803-3814.

[4] Horak, K., Kalova, I., Eyes Detection and Tracking for Monitoring Driver Vigilance, *The 33rd International Conference on Telecommunications and Signal Processing*, Baden, 2010, pp. 204-208. ISBN 978-963-88981-0-4.

[5] Nikhil, R. P., Sankar, K. P., A review on image segmentation techniques, *Pattern Recognition*, Volume 26, Issue 9, September 1993, Pages 1277-1294, ISSN 0031-3203.

[6] Batista, J., A Drowsiness and Point of Attention Monitoring System for Driver Vigilance, *Intelligent Transportation Systems Conference*, 2007, pp. 702-708.

[7] Wang, R, Guo, L., Tong, B., Jin, L., Monitoring mouth movement for driver fatigue or distraction with one camera, *The 7th International IEEE Conference on Intelligent Transportation Systems*, 2004, pp. 314- 319.

[8] Papageorgiou, C.P., Oren, M., Poggio, T., A general framework for object detection. *Sixth International Conference on Computer Vision*, 1998, pp. 555-562.

[9] Viola, P., Jones, M.J., Robust Real-Time Face Detection. *International Journal of Computer Vision*, Springer Netherlands, 2004, pp. 137–154, ISSN 0920-5691.

[10] Lienhart, R., Maydt, J., An extended set of Haar-like features for rapid object detection. *International Conference on Image Processing*, 2002, pp. I-900- I-903.

[11] Mitra, S., Yanxi, L., Local facial asymmetry for expression classification. *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004. pp. II-889- II-894.

[12] Lowe, D. G., Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, Springer Netherlands, 2004, pp. 91-110, ISSN 0920-5691.

[13] Loy, G., Eklundh, J., Detecting Symmetry and Symmetric Constellations of Features. *Lecture Notes in Computer Science*, Springer Berlin/Heidelberg, 2006, pp. 508-521.

[14] Bergasa, L.M., Nuevo, J., Sotelo, M.A., Barea, R., Lopez, M.E., Real-time system for monitoring driver vigilance. *IEEE Transactions on Intelligent Transportation Systems*, 2006, pp. 63-77.