Digital Surface Models in Urban Areas based on Satellite Imagery

Karsten JACOBSEN and Abdalla ALOBEID

Leibniz University Hannover, Institute for Photogrammetry and Geoinformation
Nienburger Str. 1, 30167 Hannover, Germany
{jacobsen, alobeid}@ipi.uni-hannover.de

Abstract. With increasing resolution of satellite imagery, object determination, which was a domain of aerial images, became possible with space images. The determination of high resolution digital surface models (DSM) in urban areas is discussed. With area based matching a good accuracy can be reached, but area based matching does not allow a satisfying determination of the building shape. Manual stereoscopic measurements are too time-consuming, so there is no discussion about the need of an automatic procedure. By pixel based matching with dynamic programming (DP) according to Birchfield and Tomasi and semiglobal matching (SGM) according to Hirschmüller, the building shape can be extracted automatically. DP has the disadvantage of striping, requiring a post processing. With the more time-consuming SGM this problem does not exist. DP and SGM are requiring epipolar images, a fast and satisfying method for the generation is shown.

Keywords: satellite imagery, pixel based matching, epipolar images, semiglobal matching

1 Introduction

Digital Surface Models (DSMs) of urban areas are becoming more and more important. The shape of buildings of course can be generated by manual stereoscopic plotting, but this is very time-consuming. So the building shapes have to be generated automatically. Aerial images may not be available or the access is restricted in several countries, in addition very high resolution satellite have a resolution overlapping with aerial images, so the use of very high resolution satellite images is an alternative. Of course for three-dimensional object determination a stereo model is required. The images of the stereo model should come from the same orbit to avoid problems caused by changed shadows. IKONOS is a satellite, which can rotate fast enough for generating stereo models from the same orbit. This still has been improved with WorldView-1 and WorldView-2 rotating very fast, but also with GeoEye-1 stereo combinations can be generated with 0.5m ground sampling distance (GSD).

The automatic extraction of DSMs in urban areas has been made with some IKONOS stereo pairs, having the standard 1m GSD resolution. Traditionally the height information was extracted by area based matching. By least squares matching satisfying height accuracy can be reached, but the least squares matching has the disadvantage of smoothening objects with sudden height changes as it is the case for buildings. The results of the simulation of a height profile determination by area based matching can be seen in figure 1. It is based on a view with 17° from left and 17° from right hand side. This corresponds to the common height to base relation of 1.6. The template size varies from 10 x 10 pixels over 8 x 8 to 6 x 6 pixels. Obviously with a smaller template size the smoothening effect is reduced, but even with 6 x 6 pixels the shape of the buildings cannot be determined very well. Operationally a small template size is not leading to acceptable results caused by the image noise, so in most cases the least squares matching needs a template size of at least 10 x 10 pixels, smoothing the DSM. With a combination of eccentric pattern matrixes the fit to buildings can be improved, allowing a correct shape determination of large buildings, but not for small buildings, building extensions and buildings with not flat roofs. An improvement is possible by feature based matching e.g. with interest operators or SIFT, but this does not lead to satisfying details.

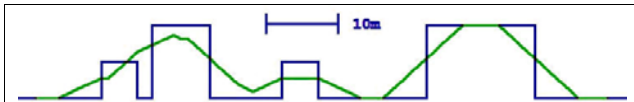
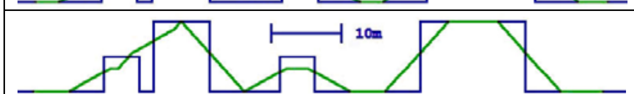
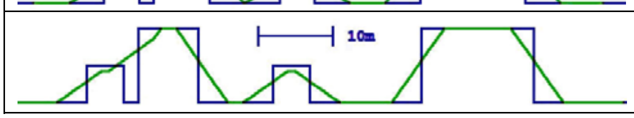
	View direction -17°, 17° Template size 10x10 pixels, 1m GSD
	View direction -17°, 17° Template size 8x8 pixels 1m GSD
	View direction -17°, 17° Template size 6x6 pixels 1m GSD

Figure 1: simulated height profiles according to area based matching (green = inclined lines = generated by area based matching)

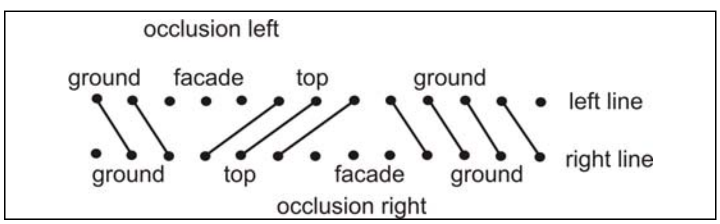


Figure 2: problem of corresponding pixels in epipolar line pair including a building – the facades are only available in one image, causing occlusions at first in left, later in right image – the building top corresponds to a sudden change of the x-parallax against the ground (3 pixels here)

2 Test Data Sets

Three IKONOS stereo scene combinations have been analyzed in urban areas. The height to base relation of 7.5 corresponds to an angle of convergence of just 7.5°, while the values 1.6 and 1.7 are corresponding to 35° respectively to 32°.

Table 1: technical data of analyzed IKONOS stereo scenes

Test area	Maras	Istanbul	San Diego
Acquisition	30.01.2003	16.05.2005	07.02.2000
Height-base-relation	7.5	1.6	1.7
Sun elevation	50.8°	65.5°	34.2°

3 Generation of Epipolar Image Pairs

The height determination by least squares matching can be made with the original IKONOS Geo scenes, allowing parallaxes in x'- and y'-direction. But the both other methods, the pixel based matching with dynamic programming and the semiglobal matching require epipolar images – images where the object heights are causing changes just in the image x'-direction or in other words, lines in both epipolar images with same y-value are corresponding to each other, including the same imaged object elements. IKONOS Geo scenes are projections of the satellite images to a plane with constant height, oriented in north direction. Such images can be transformed into epipolar scenes, just by rotating the images to the base direction. The base direction of perspective images is the direction from one projection centre to the other. In the case of line scan images as IKONOS we have not one projection centre, we have projection centre lines, but for IKONOS the field of view is just 0.9°, so the projection centre corresponding to the scene centre can be used. The Hannover satellite orientation program CORIKON computes the projection centre lines by means of the nominal collection elevation and azimuth together with the satellite orbit, allowing the determination of the base direction related to the scene centres of a stereo pair. The IKONOS Geo scenes rotated to the base direction (approximately 80°), and shifted in y-direction if the scenes do not have the same size, are with satisfying accuracy epipolar images. A comparison of corresponding image points in the epipolar images showed root mean square y-parallaxes below 1 pixel. This is satisfying for image matching searching corresponding points just in the epipolar line direction.

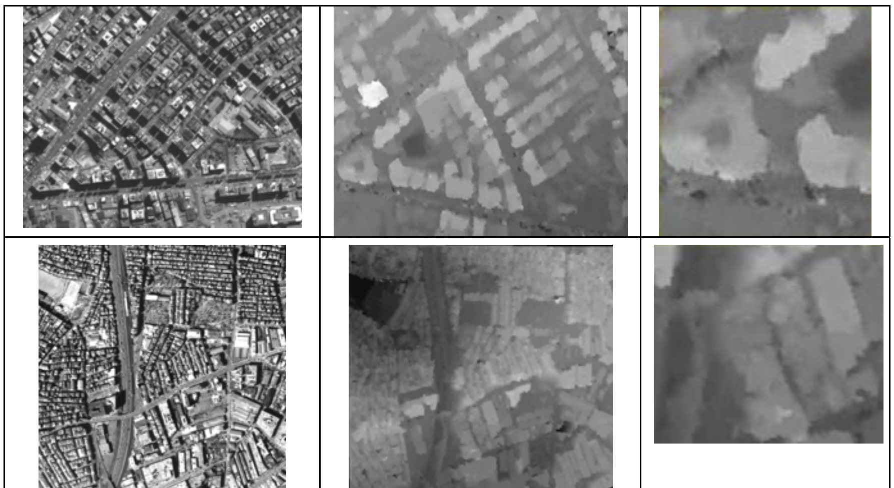
4 Matching Algorithm

The automatic image matching started with the cross-correlation, finding the corresponding image positions in the two images of a stereo pair at the position of maximal correlation coefficient of a pattern matrix in a search matrix. This type of image matching with its different versions, e.g. vertical line locus, is operational and leads to satisfying results in open and flat areas. In inclined locations the size of the sub-matrixes in both images does not cover the same size in object space, leading to a degradation of the accuracy. The least squares matching (LSM) [Förstner, 1982] avoids this problem by affine transformation of one sub-matrix to the other (formula

2). In addition to the geometric fit of the sub-matrixes, a grey value shift and grey value drift in x- and y-coordinate direction is adjusted (formula 1).

$V(x,y) = g\ (x, y) – [r_0 + r_1 * g'\ (x', y')\ ]$.	(1)
$x' = a_0 + a_1 * x + a_2 * y \qquad y' = a_4 + a_5 * x + a_6 * y$.	(2)

Least squares matching, g= grey value left, g'=grey value right, r_n, a_n = unknowns



IKONOS sub-images	Grey value coded DSM	Sub-area of grey value coded DSM

Figure 3: DSM generated by least squares matching – above Maras, below Istanbul

The results of image matching by LSM shown in figure 3 with the grey value coded digital surface models (DSMs) shows the typical blurring at objects as buildings with sudden change in height. Especially in the enlargement on right hand side it becomes obvious, that the building shape cannot be determined in a satisfying manner by area based matching confirming the simulation in figure 1.

The algorithm for pixel based matching with dynamic programming of [Birchfield, Tomasi 1999], requiring epipolar images, avoids the problem of area based matching by working pixel for pixel. It is based on the dissimilarity between the pixels in corresponding epipolar lines, identical to the absolute difference of grey values. This method requires an additional constraint. The algorithm is focused especially on the determination of height values for the case of sudden height changes, avoiding the smearing effects of LSM. Each pixel of one epipolar line is compared with all pixels of the corresponding epipolar line in the range of the maximal expected x-parallax, combined with constraints to reward successful matches and to respect occlusions. The matching is expressed as an optimization for each pair of epipolar lines, based on a pre-defined cost function. The grey value of each pixel of one line is compared with the grey values of all pixels of the other line, leading to a 2D dissimilarity matrix. As compensation of the integer nature of sampling, not the original grey values are used. In relation to the handled pixel the grey values of the average to the neighboured pixels in the epipolar line are computed. From these average values and the grey value of the handled pixel, the minimum dissimilarity to the reference grey value of the pixel in the other image is used. From the minimal differences left with ½-neighbours to right and minimal difference right with ½-neighbours to left, the minimal value finally is used. This is insensitive to sampling.

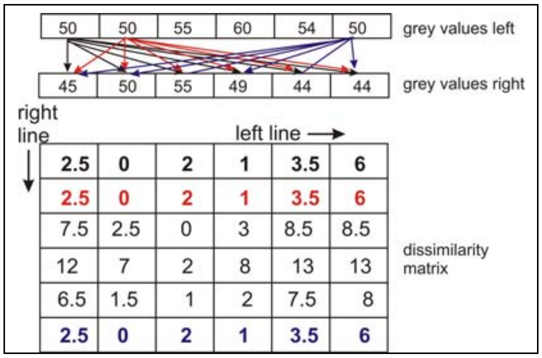


Figure 4: dissimilarity matrix based on grey values of 2 epipolar lines respecting minimum of ½-neighbours, to right = left epipolar line, down = right epipolar line

The optimal path through the dissimilarity matrix would lead to corresponding pixels, but this would cause noisy results, so based on the dissimilarity matrix a cost function with additional components is generated. The cost function λ(x,y) has the 3 components shown in formula 3:

$$\lambda(x,y) = \sum_{i=1}^{N_m} d(x_i, y_i) - N_m * K_r + N_{occ} * K_{occ}$$

(3) Cost function of DP

The first component is the sum of the dissimilarities $d(x_i,y_i)$ between the matched pixels (fig. 4), it should dominate the cost function.

The second component (N_m*K_r) is a reward for correct matching, where N_m is the number of matched pixels and K_r is the chosen match reward per pixel. As value for the empirical match reward for the handled data, values between 3 and 12 have been used. K_r is interpreted as the maximum amount of dissimilarity (difference of grey values) expected between two matching pixels.

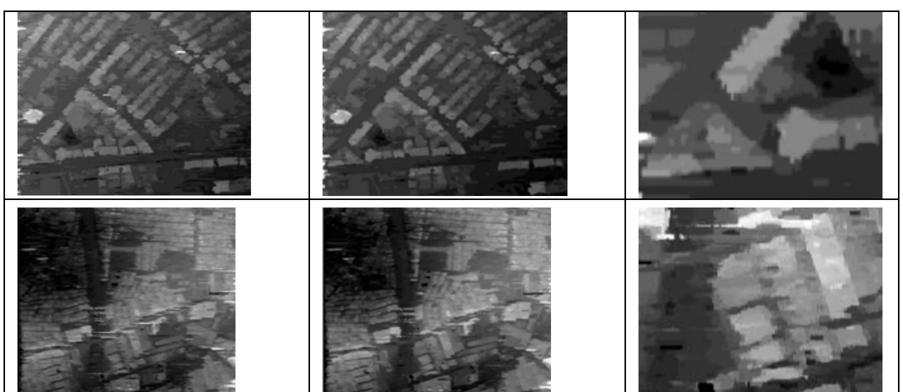
The third component ($N_{occ}*K_{occ}$) is a penalty for occlusions, where N_{occ} is the number of occlusions (not the number of occluded pixels) and K_{occ} is the chosen occlusion penalty. As value for the empirical occlusion penalty for the handled data, values between 10 and 46 have been used. The number of matched pixels plus the number of occluded pixels is identical to the number of pixels in the handled line – in the case of the example shown in figure 2, 3 pixels are occlusions.

Based on formula 3 with some side conditions a cost matrix is generated by forward-looking algorithm (fig. 5). By dynamic programming the optimal path (lowest sum of cost) in the cost matrix is determined backward in the range from the first column to the last row, giving the geometric relation between the left and the right epipolar line, as shown with the black marked optimal path in figure 3. The path doesn't have to be searched in the region exceeding the chosen limit of possible size of x-parallax (7 pixels in fig.5). The black path corresponds to a fast search, preferring the expansion to neighboured row / path with the lower value, searched from both directions.

Figure 5: dissimilarity matrix with 3 possible paths, the black path has the lowest cost, corresponding to the optimal solution (geometric relation of left to right epipolar line), no search in upper left and lower right triangle because exceeding maximal x-parallax

path	Σ dissimilarity	matched pixels	reward Kr	occlusions	Penalty Kocc	cost
M1	68	8	(-)5	5	20	128
M2	66	8	(-)5	5	20	126
M3	65	9	(-)5	4	20	100

Table 2: cost for the path in figure 5 with Kr=5 and Kocc=20



Grey value coded original DSM	Grey value coded DSM after filtering	Sub-area of grey value coded DSM, filtered

Figure 6: DSM generated by pixel based matching with dynamic programming – above Maras, below Istanbul – same area as in figure 3

The pixel based matching with dynamic programming handles any epipolar line pair individually. The neighbourhood is not respected, showing matching problems by striping in x-direction as it can be seen in figure 6, left hand side. The disturbing striping has been reduced by Median filter with a filter matrix of just 1 pixel in x-direction and 7 pixels in y-direction to avoid a degradation of the building shape. This has a similar function as the post processing propagated by [Birchfield and Tomasi, 1999] in chapter 8. A larger filter matrix leads to less clear building shapes. In general the building shapes are quite clearer as achieved by least squares matching.

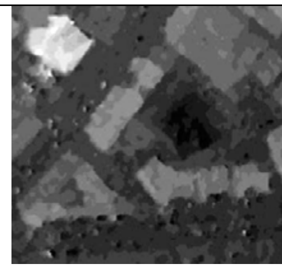
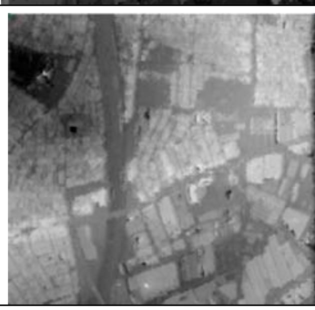
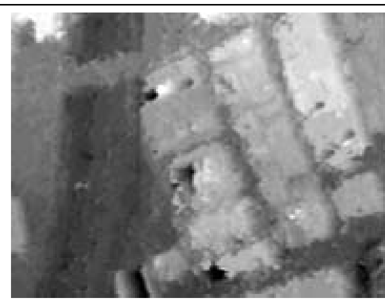
		IKONOS Maras
		IKONOS Istanbul
Grey value coded DSM	Sub-area of grey value coded DSM	

Figure 4: DSM generated by Semiglobal Matching, same area as in figure 2

Semiglobal matching (SGM) [Hirschmüller 2008] is an extension of the DP, it avoids the problem that any pair of epipolar lines is handled separately, leading to striping. First, the dissimilarity is expressed pixel per pixel by Mutual Information (MI). MI finds the correspondence between conjugate pixels without assuming that conjugate points have similar intensity values. Instead, the joint probability distribution in the form of the joint intensity value histogram is used. MI has been shown to be robust in respect to radiometric differences. A global 2D smoothness constraint across multiple intersecting lines is introduced. It is approximated by combining many one-dimensional constraints. The first step for SGM is to obtain an initial disparity matrix that is required for warping one of the stereo images before MI can be calculated. Corresponding to [Hirschmüller 2008] the process is started with a random disparity image, and then continued in a hierarchical manner. Subsequently, the joint histogram is derived over the whole images. It is stored as a 256×256 histogram and is smoothed using a Gaussian kernel. Then, the MI values are computed. The third step is the determination of a disparity image that minimizes the energy function by path wise optimization of several one-dimensional-paths toward the pixel under consideration. After this, the costs are summed up over all paths in relation to the handled pixel. The final disparity image is computed based on the minimal cost.

A visual inspection of the grey value coded DSMs in urban areas (figure 4) shows the clear improvement against DP. No striping can be seen and the building shapes became clear. Also smaller objects, as parking cars, became very clear. Such small objects have been eliminated by the required Median filter following pixel based matching with dynamic programming. The cars are still visible in the height models based on least squares matching, but they are not as clear as with SGM.

5 Analysis of achieved results

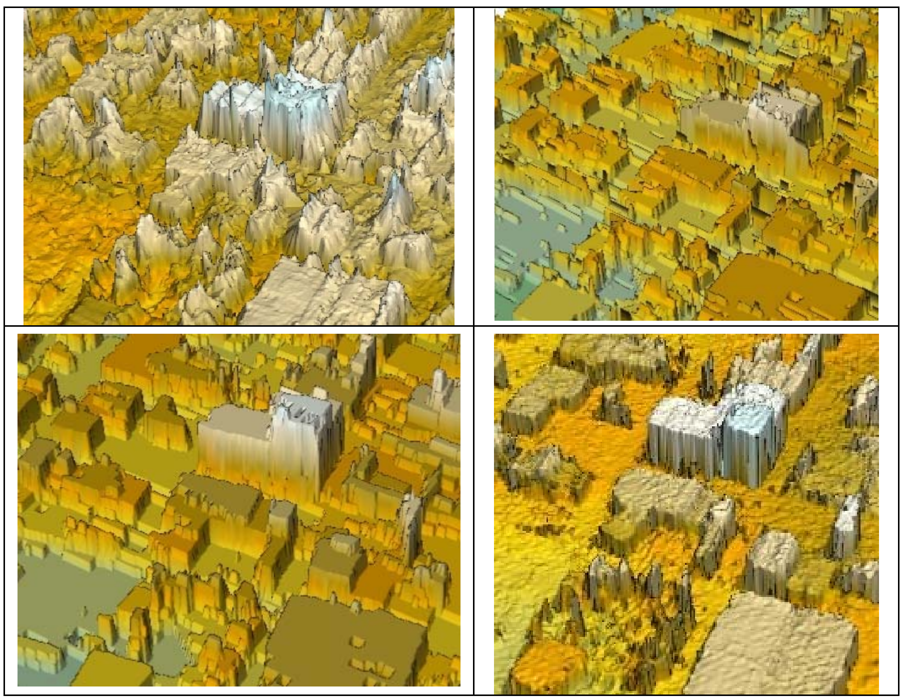


Figure 5: 3D-view to generated digital surface models, IKONOS San Diego
above: LSM above: DP without filtering
below: DP with filtering below: SGM

Fig. 6: profiles through DSMs, left: LSM, centre: DP without filter, right: SGM

The 3D-views to the digital surface models generated by the described matching algorithm (figure 5) and the profiles in figure 6 shows clearer as the grey value coded DSMs the advantages and disadvantages of the methods. The shape of the buildings determined by LSM is not very clear, the facades are tilted. The DP without filtering is influenced by striping, causing disturbances looking like walls – partially crossing streets. Most of these effects have been removed by Median filtering just across the epipolar lines. But with SGM such effects do not exist and also the fine structure on top of buildings is more realistic. Nevertheless, also SGM had some problems with a group of smaller buildings in the lower left side of figure 5. This group of buildings is more realistic in the DSM based on filtered DP. The filtered DP is generalizing the results, which sometimes is requested. The profiles through the height models (figure 6) confirm the findings – the building shapes determined by LSM are not clear, the facades are inclined and the building tops are not flat. If the DSM determined by DP is not filtered, it shows several spikes. The building shapes from SGM are clear and realistic.

		LSM	**DP**	**SGM**
Maras	height-base-relation = 7.5	0.9m	1.1m	0.5m
Istanbul	height-base-relation = 1.6	0.7m	0.8m	0.6m
San Diego	height-base-relation = 1.7	0.8m	1.0m	0.4m

Table 3: standard deviation of building heights by different matching algorithm

The visual inspection is only one part of the analysis. The matching results also have been geometric verified. In the San Diego test area the building heights have been measured stereoscopically by an experienced operator. In Istanbul reference heights from stereoscopic measurement of aerial images have been available and in Maras the building heights have been estimated by the length of building shadows on flat terrain together with the sun elevation available in the IKONOS header files.

The standard deviation of the building heights shown in table 3 is without the influence of the accuracy of the reference measurements, which has been respected. In all three data sets the results achieved with LSM is slightly better as with DP. For the determination of the building height, the centre of the building and not the edges smoothened by LSM, have been used, improving the results achieved with LSM. DP is based just on one vector of grey values and not as the LSM on a matrix of grey values. By this reason the DP is noisier, reducing slightly the accuracy. The SGM is not influenced by smoothing effects and because of the used high number of profiles, the results are based on an area, reducing noise - this explains the clearly better results achieved by SGM. The accuracy achieved in all three test areas is similar, even if the height to base relation of the test area Maras is very large. This confirms earlier results [Büyüksalih, Jacobsen 2007], that the height to base relation is not so important because the disadvantage of a smaller angle of convergence is at least partially compensated by better matching results in the image space caused by more similar images.

6 Conclusion

The automatic determination of building shapes based on a very high resolution stereoscopic stereo pair became possible with automatic image matching by pixel based matching with dynamic programming according to [Birchfield & Tomasi 1999] and with Semiglobal Matching [Hirschmüller 2008]. With traditional area based matching of Least Squares Matching [Förstner 1982], the building outlines cannot be determined in a satisfying manner. DP has the disadvantage of not respecting the neighboured lines, leading to a striping of the determined DSM, but with Median Filter just across the epipolar line the striping can nearly be eliminated. The Median filter has a similar effect as the post-processing described by [Birchfield & Tomasi 1999]. The filtering causes a loss of small details, so the buildings do not show small elements on the top of the buildings. This corresponds to a generalization, which sometimes should be made. The more computing intensive SGM describes the visible surface more in detail, but it also may fail in case of too small structures. In general with both methods of pixel based matching the automatic generation of building shapes became possible.

References

1. Alobeid, A., Jacobsen, K., Heipke, C., 2009: Building height estimation in urban areas from very high resolution satellite stereo images, ISPRS Hannover Workshop 2009, IntArchPhRS. Vol XXXVIII-1-4-7/W5
2. Birchfield, S., Tomasi, C., 1998: A pixel dissimilarity measure that is insensitive to image sampling. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(4):401-406
3. Birchfield, S., Tomasi, C., 1999: Depth discontinuities by pixel-to pixel stereo. International Journal of Computer Vision 35(3): 269-293
4. Büyüksalih, G.; Jacobsen, K., 2007: Digital Surface Models in Build up Areas based on very high resolution Space Images: ASPRS annual conference. Tampa, on CD, also available at www.ipi.uni-hannover.de (last access April 2009)
5. Förstner, W., 1982: On the geometric precision of digital correlation, IntArchPhRS, (24)3, pp. 176- 189
6. Hirschmüller, H., 2008: Stereo Processing by Semiglobal Matching and Mutual Information, IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(2):328-34