

WO 94/17485

PCT/FR94/00057

1

PROCESS FOR ESTIMATING DISPARITY BETWEEN THE MONOSCOPIC  
IMAGES MAKING UP A STEREOSCOPIC IMAGE

The present invention relates to a process for  
5 estimating disparity between the monoscopic images  
making up a stereoscopic image. It applies in  
particular to the processing of moving stereoscopic  
images digitized in sequences, this processing making  
it possible to obtain a description of the relief  
10 through measurements of the distances from the movie  
cameras. It applies for example to the processing of  
the signal from stereoscopic televisions or to the  
special effects processing of a stereoscopic image.

Stereoscopic image processing, for example  
15 making an object stand out from any type of background,  
other than blue ones in particular, requires a  
knowledge of the relief of this object.

A stereoscopic image results from the combining  
of two monoscopic views captured by two cameras placed  
20 at different spots, the distance between the cameras  
being for example equal to the distance between two  
eyes and their direction simulating the human gaze for  
example. Each stereoscopic image therefore consists of  
a pair of monoscopic images. It is known that a  
25 knowledge of the point-to-point disparity between the  
images of one and the same pair makes it possible to  
deduce the distance of the filmed objects from the  
cameras and thus to deduce therefrom a relief datum for  
these objects.

30 A physical description of the differences  
between the two images of a stereoscopic pair is  
obtained through point-to-point matching providing  
vector data. The disparity field consists of these  
vector data which are in fact related to the change of  
35 perspective between the two images and to the distance  
from the cameras to the objects of the scene to be  
filmed.

A method of point-to-point matching of two  
images, a left image and a right image, making up a

THIS SHEET HAS BEEN  
REPLACED



stereoscopic image is described in an article:  
PROCEEDINGS OF THE CONFERENCE ON COMPUTER VISION AND  
PATTERN RECOGNITION, Miami Beach, 22-26 June 1986,  
pages 219-226, IEEE, Silver Spring, US; L.R. WILLIAMS  
5 et al.: "A coarse-to-fine control strategy for stereo  
and motion on a mesh-connected commuter".

A hierarchical decomposition of the images is  
carried out there by levels of resolution, and then an  
estimate of the disparities between these images is  
10 made, starting from the coarsest level of resolution  
and going towards finer and finer resolutions. This  
method, using a single pyramid of images of increasing  
resolutions, does not always afford a sufficiently  
accurate estimate, especially for slender objects or  
15 for outlines, or again when the regions of disparities  
are substantially uniform.

The point-to-point disparities between the two  
images of one and the same pair can be described by  
vectors, called disparity vectors, the purpose of the  
20 invention is to obtain a sequence of fields of  
disparity vectors corresponding to a sequence of  
stereoscopic images satisfactorily estimating the  
differences between the two views of a stereoscopic  
image, especially with a high accuracy.

25 To this end, the subject of the invention is a  
process for estimating disparity between the monoscopic  
images making up a stereoscopic image, characterized in  
that it consists:

- in a first step, in carrying out one and the same  
30 hierarchical decomposition of each monoscopic image, by  
levels of resolution, thereby building at least three  
pyramids of images of increasing resolutions, one low-  
frequency image pyramid and two high-frequency image  
pyramids in the horizontal and vertical directions of  
35 the images;

- in a second step, in estimating a disparity at the  
coarsest level of resolution and in refining the  
estimate in each level as the resolution increases, the  
disparity being represented at each point of the image



by a vector whose origin is situated at the point of one of the monoscopic images and on the tip at the corresponding point of the other monoscopic image.

The main advantages of the invention are that  
5 it improves the accuracy of estimation, especially as regards outlines or slender objects, that it can be used for numerous applications in stereoscopic image processing, and that it makes it possible to associate  
10 estimation of motions of objects of a stereoscopic image with the estimation of disparity and that it is simple to implement.

Other characteristics and advantages of the invention will emerge with the aid of the description which follows, given in connection with the appended  
15 drawings which represent:

- Figure 1, a pair of images constituting a stereoscopic image, which are obtained with two cameras filming the same scene at two different angles;
- Figure 2, the preceding images placed side by side;
- 20 - Figure 3, a sequence of stereoscopic images;
- Figure 4, two series of successive left and right images of a sequence of change of a stereoscopic image;
- Figure 5, a pyramidal hierarchical image decomposition;
- 25 - Figure 6, illustrations of possible predictions;
- Figure 7, illustrations of possible cost functions.

A stereoscopic pair of images can be modelled by two perspective projections in different image planes of the same three-dimensional scene. Figure 1  
30 depicts a pair of natural, monoscopic images  $I_l(t)$ ,  $I_r(t)$ , a left image  $I_l(t)$  and a right image  $I_r(t)$  captured at an instant  $t$ , and obtained for example with the aid of two cameras filming the same scene from two different angles.

35 The problem of estimating disparity consists in the point-to-point matching of the two images or views, right and left. If a point  $P$  in space, with spatial coordinates  $(X, Y, Z)$  is projected respectively onto the left image  $I_l(t)$  and onto the right image  $I_r(t)$  at



a point  $Z_l$  with coordinates  $(X_l, Y_l)$  and at a point  $Z_r$  with coordinates  $(X_r, Y_r)$ , the disparity vector  $\bar{D}$  assigned to the point  $Z_r$  of the right image has components  $(X_r - X_l, Y_r - Y_l)$ . It is known that the  
5 disparity datum makes it possible to calculate the distance from the points of the filmed scene to the camera, once the optical and geometrical parameters of the acquiring of the images are known. These parameters are, in particular, the focal lengths, the optical  
10 centres and the distance between the two cameras.

Figure 2 depicts, side by side, the two monoscopic images  $I_l(t)$ ,  $I_r(t)$  of the preceding figure. A disparity vector at an instant  $t$ , is defined at each point  $Z_r$  of the right image  $I_r(t)$ , the point  $Z_r$  being  
15 labelled by a pair of coordinates  $(x, y)$  in this image  $I_r(t)$ . The disparity vector at this point  $Z_r$  is for example denoted  $\bar{D}(Z_r)$ . The point of the left image  $I_l(t)$  corresponding to the point  $Z_r$  of the right image  $I_r(t)$  being denoted  $Z_l$ , it is connected with the first  
20 point  $Z_r$  through the following relation:

$$Z_l = Z_r - \bar{D}(Z_r) \quad (1)$$

A point  $Z'_l$  of the right image  $I_r(t)$  represents the point  $Z_l$  of the left image  $I_l(t)$  in the latter. It is also connected with the point  $Z_r$  of the right image  
25  $I_r(t)$  through relation (1).

According to the invention, the process for estimating stereoscopic disparity between the left image  $I_l(t)$  and the right image  $I_r(t)$  uses a technique for estimating motion between two consecutive images of  
30 a monoscopic sequence described in French Patent No. 2,663,178: "Procédé d'estimation hiérarchique du mouvement dans une séquence d'images", [Process of hierarchical estimation of motion in an image sequence], the right image  $I_r(t)$  being akin to an image  
35 consecutive to the left image  $I_l(t)$  and the disparity vector  $\bar{D}$  being akin to the motion vector. Thus, in a first step, one and the same hierarchical decomposition of each monoscopic image  $I_r(t)$ ,  $I_l(t)$  is carried out by levels of resolution. To do this, at least three



pyramids of images of increasing resolution are constructed: one low-frequency image pyramid and two high-frequency image pyramids in the horizontal and vertical directions of the images  $I_r(t)$ ,  $I_l(t)$ . In a second step, a disparity is estimated at the coarsest level of resolution, then this estimate is refined in each level of resolution as the resolution increases.

The disparity is analysed with the aid of a hierarchical decomposition of the monoscopic images analysed  $I_r(t)$ ,  $I_l(t)$  into sub-bands, while preserving at each level of resolution of the decomposition the low-frequency band and the high-frequency bands respectively representing the vertical and horizontal outlines of the image.

Figure 3 depicts a sequence of stereoscopic images which is made of left monoscopic images  $I_l$  and right monoscopic images  $I_r$  changing as a function of time  $t$ .

With the stereoscopic image changing over time, as Figure 3 illustrates, it may be necessary to make temporal predictions of the disparities between the left images  $I_l$  and right images  $I_r$  of the stereoscopic image.

Figure 4 depicts two series of successive left and right images of a sequence of change of a stereoscopic image. The left monoscopic image  $I_l(t)$  captured at an instant  $t$  is preceded by a left monoscopic image  $I_l(t-1)$  captured at an instant  $t-1$ . Similarly, the right monoscopic image  $I_r(t)$  captured at an instant  $t$  is preceded by a right monoscopic image  $I_r(t-1)$  captured at an instant  $t-1$ .

The temporal prediction of the disparity vector  $\vec{D}(t)$  at a point of the right image  $I_r(t)$  is calculated from the disparity vector  $\vec{D}(t-1)$  corresponding to this point in the preceding right image  $I_r(t-1)$ , with the aid of the motion vectors calculated between the instant  $t$  and the preceding instant  $t-1$  on the right  $I_r(t-1)$ ,  $I_r(t)$  and left  $I_l(t-1)$ ,  $I_l(t)$  images. These vectors are denoted  $V_r(t)$  and  $V_l(t)$  respectively,

called the right motion vector and left motion vector.

The aforesaid disparity and motion vectors are connected by the following relation:

$$\bar{D}(t) + \bar{V}l(t) = \bar{D}(t-1) + \bar{V}r(t) \quad (2)$$

In practice, the temporal prediction of the disparity vector  $\bar{D}(t)$  at a current point  $Z_{rc}$  of the right image  $I_r(t)$  at an instant  $t$  can for example be obtained as follows: the corresponding point  $Z_{rp}$  in the right image  $I_r(t-1)$  at the preceding instant  $t-1$  is known through the right motion vector  $\bar{V}r(t)$  at the instant  $t$ , hence these two points  $Z_{rc}$ ,  $Z_{rp}$  are connected by the following relation:

$$Z_{rp} = Z_{rc} - \bar{V}r(t) \quad (3)$$

The point  $Z_{lp}$  of the left image  $I_l(t-1)$  at the preceding instant  $t-1$  corresponding to the point  $Z_{rp}$  of the right image  $I_r(t-1)$  captured at the same instant  $t-1$  which corresponds to the current point  $Z_{rc}$  of the right image  $I_r(t)$  captured at the instant  $t$  is known through the disparity vector  $\bar{D}(t-1)$  of the preceding instant  $t-1$  in accordance with the following relation:

$$Z_{lp} = Z_{rp} - \bar{D}(t-1) \quad (4)$$

The point  $Z_{lc}$  corresponding to this point  $Z_{lp}$  in the left image  $I_l(t)$  at the instant  $t$  can be obtained with the aid of the left motion vector  $\bar{V}l(t-1)$ , captured at the preceding instant in accordance with the following relation:

$$Z_{lc} = Z_{lp} + \bar{V}l(t-1) \quad [sic] \quad (5)$$

A temporal prediction of the disparity vector  $\bar{D}(t)$  at the current point  $Z_{rc}$  of the right image  $I_r(t)$  at the instant  $t$  is given by the vector formed from the pair of points  $(Z_{lc}, Z_{rc})$  where  $Z_{lc}$ , corresponding to the current point  $Z_{rc}$  of the right image  $I_r(t)$  in the left image  $I_l(t)$  at the same instant  $t$  is calculated with the aid of the last three relations (3), (4), (5) established earlier. In this last case, an estimate of motion and of disparity is thus proposed, in which the disparity is predicted temporally by itself, while correcting this prediction by any possible motion of the objects in the stereoscopic scene.



A joint estimate of disparity and of motion can be produced, the motion is estimated for example independently over each of the two right Ir and left Il image sequences. In this case, the motion vector  
5 calculated over one sequence, right or left, is predicted temporally only by the motion vectors calculated previously over the same sequence. The motion estimation technique can be that described by French Patent No. 2,663,178. The disparity is, for its  
10 part, predicted temporally by the disparity calculated at the preceding instants, compensated by the motions calculated, independently, over the right Ir and left Il images. Relations (3), (4), (5) show that the prediction vector  $\bar{D}(t)$  at an instant t requires a  
15 knowledge of the prediction vector  $\bar{D}(t-1)$  at the preceding instant as well as of the right motion vector  $V_r(t)$  at the instant t and of the left motion vector  $V_l(t-1)$  at the instant t-1.

Figures 5 to 7 accompany the description of  
20 improvements to the estimation process described in French Patent No. 2,663,178, which are applicable to the motion estimation or disparity estimation of the present invention. The description of the aforesaid improvements is illustrated by its application to  
25 disparity estimation.

As illustrated in Figure 5, the source images 51, 52, 53, 54 of luminances are decomposed into a multiresolution pyramid. The source image, possessing a number L of lines and a number C of columns, is by  
30 definition the image 51 of level 0, with resolution  $1/2^0 = 1/1$  in such a pyramid. In this pyramid an image of level i, with resolution  $1/2^i$ , is made up of  $L/2^i$  lines and  $C/2^i$  columns. It is obtained through i low-pass filterings and i successive linewise and  
35 columnwise sub-samplings by 2. This low-frequency image pyramid is deduced from a hierarchical decomposition into sub-bands such as described by French Patent No. 2,663,178.

Image analysis takes place hierarchically over



the pyramid. It then consists in firstly making a global estimate at a coarse level of resolution of the image, and in then refining this estimate as the resolution of the processed images increases. At each level of resolution  $1/2^i$  in the pyramid, the disparity vectors estimated at the immediately coarser level of resolution  $1/2^{i+1}$  have their moduli multiplied by two in order to take into account the sub-sampling factor and are used for prediction. The aforesaid patent describes how this prediction can be improved by proposing four predictions, rather than just a single one arising from an immediately preceding coarser level, for the calculation of a disparity vector at the current level. This amounts to including the contribution from a vector of the coarse level for the prediction of four to sixteen points of the following finer level. However, it is possible for example to reduce the number of predictions arising from the coarse level to three for the calculation of a vector of the fine level. This amounts to including the contribution from a vector of the coarse level for the prediction of twelve points only of the fine level as Figure 6 illustrates. In this figure, a prediction is illustrated by an arrow 61 connecting a coarse level 62 to a finer level 63.

The aforesaid predictions 61 can each be improved by a further temporal prediction. Thus, a disparity vector estimated at a given point at a preceding instant  $t-1$  between right  $I_r(t-1)$  and left  $I_l(t-1)$  images at the same level of resolution is for example projected temporally in the direction of the motion vector associated with the given point towards the images  $I_r(t)$ ,  $I_l(t)$  captured at the instant  $t$ . A method of temporal prediction which can be applied is described in French Patent No. 2,675,002: "Procédé de classification des pixels d'une image appartenant à une séquence d'images animées et procédé d'interpolation temporelle d'images utilisant ladite classification" [Process for the classification of the pixels of an



image belonging to a sequence of moving images and process for the temporal interpolation of images using the said classification].

Thus, the calculation of the disparity vector is performed at every point of the image, level after level in the multiresolution pyramid, from the coarsest to the finest constituted by the original image which is not sub-sampled. Four prediction vectors can be proposed when calculating each disparity vector: one vector arising from the vector calculated at the preceding instant, projected in the direction of motion, and three calculated at the immediately coarser level. It follows that only a small correction is necessary around such a prediction in order to obtain the correct disparity vector.

A disparity vector to be estimated  $\bar{D}$  is the sum of a prediction vector  $\bar{P}$  and a correction  $\bar{\delta}$ , i.e.:

$$\bar{D} = \bar{P} + \bar{\delta} \quad (6)$$

Depending on the maximum amplitude, horizontally and vertically, of the disparity vector to be estimated, the search around the predication [sic] of the latter, represented by a vector, can for example be performed solely horizontally or in both directions, horizontal and vertical. French Patent No. 2,663,178 describes a search in both directions. Such a correction of resolution of  $\pm 1$  pixel, initialized to 0 at a level N-1 of the pyramid for example, leads to a maximum value equal to:

$$|D_{\max}| = \sum_{k=0}^{N-1} 2^k \quad (7)$$

Thus, the number of levels  $N_{\max}+1$  required to estimate a disparity with maximum amplitude equal to the aforesaid maximum value  $|D_{\max}|$  is given by the integer part of the logarithm to the base 2 of this maximum value  $|D_{\max}|$ , i.e.:

$$N_{\max} = \text{integer part of } \log_2 (|D_{\max}|) \quad (8)$$

In order, in particular, to economize on the number of levels of resolution, it is possible to separate the calculation of the disparity vector



horizontally and vertically. Let  $|Dx \max|$  and  $|Dy \max|$  be the maximum values of the disparity, horizontally and vertically respectively. A  $\pm 1$  pixel search initialized at 0 is necessary only starting from the level of order  $Nx \max$  horizontally and of order  $Ny \max$  vertically, these levels  $Nx \max$ ,  $Ny \max$  being calculated from the two preceding relations (7) and (8).  $|Dx \max|$  being for example assumed to be greater than  $|Dy \max|$ , a solely horizontal search is conducted in the pyramid between the levels  $Nx \max$  and  $Ny \max + 1$ , followed by a two-dimensional search, horizontally and vertically, between the level  $Ny \max$  and 0.

Thus, the starting level for the estimation of disparity in the pyramid is calculated as a function of the maximum disparity to be attained, fixed a priori. This calculation is for example carried out separately for the horizontal disparity and vertical disparity, a simple horizontal search being conducted for as long as the maximum amplitude, vertically, does not justify a two-dimensional search.

The disparity vector chosen is the one which minimizes a certain cost function such as described for example in French Patent No. 2,663,178. A cost function can for example be the sum of the squares of the differences in luminance between the right and left images, the sum being taken for example over a window of  $3 \times 3$  pixels, centred at the point of calculation of the disparity vectors.

It is possible to deduce a measure of confidence for the chosen disparity vector. In a homogeneous region of the image in particular, the values of the cost function being substantially equal whatever vector is tested, the confidence in the vector chosen will be low. On the other hand, in a zone of the image with a steep gradient, the correct vector corresponds to a very sharp minimum in the cost function and is therefore very reliable.

The confidence is for example calculated

separately through the horizontal component and vertical component of the vector. If  $P_x$  is the horizontal component of the prediction vector, values  $P_{x \pm n}$  are proposed for the corrected vector,  $n$  being  
 5 equal for example to 2 if the search is one-dimensional or to 1 if the search is two-dimensional, four vectors then being proposed. The vector which minimizes the cost function having been chosen, it has horizontal component  $P_{xc}$  where  $P_{xc}$  belongs to the interval  $[P_{x-n}, P_{x+n}]$ . In order to bracket the minimum value of the  
 10 cost function, the confidence is calculated over the interval  $[P_{x-n-1}, P_{x+n+1}]$ , even if the choice of minimum is made in the interval  $[P_{x-n}, P_{x+n}]$ .

With  $F$  representing the cost function, three  
 15 values of this function are available  $F(P_{xc})$ ,  $F(P_{xc-1})$  and  $F(P_{xc+1})$ . The confidence  $\text{conf}(P_{xc})$  in the horizontal component  $P_{xc}$  of the chosen vector is given for example by the following two relations:

$$20 \quad \text{Curv}(P_{xc}) = \frac{F(P_{xc-1}) + F(P_{xc+1}) - 2F(P_{xc})}{2} \quad (9)$$

$$\text{Conf}(P_{xc}) = \frac{\text{curv}(P_{xc})}{K + F(P_{xc})} \quad (10),$$

This Sheet Has Been Replaced



$K$  being a positive constant.

25 A confidence thus defined varies from 0 to  $+\infty$  and it can be advantageous to restrict its entire variations to the interval  $[0, 1]$ . To do this, it is possible to define a normalized confidence  $\text{cf}(P_{xc})$  for the horizontal component  $P_{xc}$  of the prediction vector,  
 30 this normalized confidence being defined by the following relation:

$$\text{cf}(P_{xc}) = \frac{\text{conf}(P_{xc})}{1 + \text{conf}(P_{xc})} \quad (11)$$

35 by combining relations (10) and (11), it follows that:

THIS SHEET HAS BEEN  
REPLACED

$$cf(Pxc) = \frac{curv(Pxc)}{K + curv(Pxc) + F(Pxc)} \quad (12)$$



The normalized confidence thus defined tends  
 5 asymptotically to 1 as the curvature of the curve of  
 the cost function  $F$  increases. The constant  $K$  makes it  
 possible to control the variation in confidence.

Figure 7 depicts by way of example two curves  
 of possible cost function  $F$ . The abscissa axis  
 10 represents the values of the horizontal component  $P_x$  of  
 the prediction vector and the ordinate axis represents  
 the cost function  $F$ . One curve 71 possesses a high  
 radius of curvature for the component  $P_{xc}$  signifying  
 that the chosen prediction vector possesses high  
 15 confidence. One curve 72 possesses a low radius of  
 curvature for this component, signifying that the  
 chosen vector possesses low confidence.

Relations analogous to the preceding relations  
 (9), (10), (11), (12) can be obtained for the vertical  
 20 component  $P_y$  of the chosen prediction vector, its  
 processing being identical.

A confidence measure, lying between 0 and 1, is  
 therefore associated with the horizontal component and  
 with the vertical component of each estimated disparity  
 25 vector. A corrected vector is associated with each of  
 the proposed vectors, four in number for example, and a  
 confidence and a cost value are associated with each  
 corrected vector. The corrected vector chosen is for  
 example the one which minimizes the cost function,  
 30 which therefore has the minimum cost value.

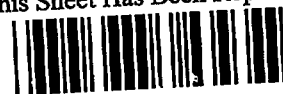
If this minimum is not sufficiently marked, the  
 choice is for example guided by the confidence.

An oriented smoothing, or filtering, of the  
 disparity vectors is for example carried out in the  
 35 multiresolution pyramid, at all levels of resolution,  
 between the estimation and prediction of the finer  
 level. This smoothing makes it possible in particular  
 to propagate the values with high confidence towards

the zones of the image with low confidence. A smoothing technique is described in particular in the article by P. Anandan: "A computational framework and an algorithm for the measurement of visual motion", International  
5 Journal of Computer Vision, 2, 283-310, 1989.

The smoothing is for example conducted independently on the horizontal component and vertical component of a disparity vector.

This Sheet Has Been Replaced



THIS SHEET HAS BEEN  
REPLACED



WO 94/17485

PCT/FR94/00057

14

## CLAIMS

1. Process for estimating disparity between the monoscopic images ( $I_r$ ,  $I_l$ ) making up a stereoscopic image, characterized in that it consists:
  - 5 - in a first step, in carrying out one and the same hierarchical decomposition of each monoscopic image ( $I_r$ ,  $I_l$ ), by levels of resolution, thereby building at least three pyramids of images of increasing resolutions, one low-frequency image pyramid and two  
10 high-frequency image pyramids in the horizontal and vertical directions of the images;  
- in a second step, in estimating a disparity at the coarsest level of resolution and in refining the estimate in each level as the resolution increases, the  
15 disparity being represented at each point of the image by a vector  $\vec{D}$  whose origin is situated at the point ( $Z_l$ ) of one of the monoscopic images and on the tip at the corresponding point ( $Z_r$ ) of the other monoscopic image.
- 20 2. Process according to Claim 1, characterized in that it consists in analysing the disparity with the aid of a hierarchical decomposition of the monoscopic images into sub-bands while preserving at each level of the decomposition the low-frequency band and the high-  
25 frequency bands respectively representing the vertical and horizontal outlines of the images.
3. Process according to any one of the preceding claims, characterized in that three prediction vectors ( $\vec{p}_1$ ) arising from the disparity vector calculated at  
30 the coarsest level are proposed as initialization of the disparity vector at the next level.
4. Process according to Claim 3, characterized in that at every level of resolution, the disparity vector ( $\vec{D}$ ) estimated at a preceding instant ( $t-1$ ) is projected  
35 temporally in the direction of the motion so as to provide a fourth prediction vector for calculation of the disparity vector.
5. Process according to any one of the preceding claims, characterized in that the starting level for

THIS SHEET HAS BEEN  
REPLACED



estimation in the pyramid is calculated as a function of a maximum disparity to be attained, fixed a priori, the calculation being undertaken separately for the horizontal disparity and the vertical disparity of the image.

6. Process according to Claim 5, characterized in that a simple horizontal search is conducted for as long as the vertical amplitude of the disparity has not attained the maximum disparity fixed a priori.
7. Process according to any one of the preceding claims, characterized in that a cost function (F) is used to choose between the estimated disparity vectors, the cost function being the sum of the squares of the differences in luminance over the points of a window centred at the point of calculation of the disparity vectors.
8. Process according to any one of the preceding claims, characterized in that a confidence function, calculated for each component of the estimated disparity vectors, is used to choose between the latter.
9. Process according to any one of the preceding claims, characterized in that at each level of resolution, after the estimation and before the prediction of each disparity vector, the latter is filtered by a confidence function.

This Sheet Has Been Replaced



THIS SHEET HAS BEEN  
REPLACED