A method for semi-automatic grading of human blastocyst microscope images

E. Santos Filho1,*, J.A. Noble1, M. Poli2,4, T. Griffiths2, G. Emerson3, and D. Wells4

1Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, Oxford OX3 7DQ, UK. 2Oxford Fertility Unit, Institute of Reproductive Sciences, Oxford Business Park North, Oxford OX4 2HW, UK. 3Human Assisted Reproduction Ireland, Rotunda Hospital, Dublin 1, Ireland. 4Nuffield Department of Obstetrics and Gynaecology, University of Oxford, Oxford OX3 9DU, UK.

*Correspondence address. E-mail: santfh@googlemail.com

Submitted on August 12, 2011; resubmitted on April 16, 2012; accepted on May 21, 2012

BACKGROUND: The precise assessment of embryo viability is an extremely important factor for the optimization of IVF treatments. In order to assess embryo viability, several embryo scoring systems have been developed. However, they rely mostly on a subjective visual analysis of embryo morphological features and thus are subject to inter- and intra-observer variation. In this paper, we propose a method for image segmentation (the dividing of an image into its meaningful constituent regions) and classification of human blastocyst images with the aim of automating embryo grading.

METHODS: The delineation of the boundaries (segmentation) of the zona pellucida, trophoderm (TE) and inner cell mass (ICM) were performed using advanced image analysis techniques (level set, phase congruency and fitting of ellipse methods). The fractal dimension and mean thickness of TE and ICM image texture descriptors (texture spectrum and grey-level run lengths) were calculated to characterize the main morphological features of the blastocyst with the aim of automatic grading using Support Vector Machine classifiers.

RESULTS: The fractal dimension calculated from the delineated TE boundary provided a good indication of cell number (presented a 0.81 Pearson correlation coefficient with the number of cells), a feature closely associated with blastocyst quality. The classifiers showed different accuracy levels for each grade. They presented accuracy ranges from 0.67 to 0.92 for the embryo development classification, 0.67–0.82 for the ICM classification and 0.53–0.92 for the TE classification. The value 0.92 was the highest accuracy achieved in the tests with 73 blastocysts.

CONCLUSIONS: Semi-automatic grading of human blastocysts by a computer is feasible and may offer a more precise comparison of embryos, reducing subjectivity and allowing embryos with apparently identical morphological scores to be distinguished.

Key words: image segmentation / computerized image analysis / embryo grading / IVF

Introduction

The process of IVF has evolved considerably since the first successful treatment three decades ago. However, the efficiency remains relatively poor, with most patients requiring more than one treatment cycle to obtain a birth. This is partly related to the great variability in developmental competence of the embryos produced during an IVF cycle and difficulties in determining which of the embryos generated has the highest probability of producing a child. For this reason, IVF clinics in many parts of the world often transfer more than one embryo per cycle, hoping to increase the odds that a viable embryo will be transferred. While this approach has helped to maintain the IVF pregnancy rates at an acceptable level, it has also led to large numbers of multiple pregnancies. The potentially negative consequences of multiple pregnancy are well documented, associated with elevated risks of serious complications, including pre-eclampsia, maternal haemorrhage, operative delivery, uterine rupture and preterm labour (Bromer and Seli, 2008).

Multiple pregnancies can easily be prevented by transferring fewer embryos to the mother’s uterus each cycle, the ideal strategy being single embryo transfer. However, restricting the number of embryos transferred risks reducing the pregnancy rates per cycle, as the probability of transferring a viable embryo by chance is diminished. This problem is most pronounced in cases of single embryo transfer. It is
essential that the embryo chosen for transfer in single embryo transfer cycles is that having the greatest potential for forming a pregnancy and producing a healthy child.

Currently, the decision of which embryo to transfer is made on the basis of morphological assessments conducted in the IVF laboratory (Cohen et al., 1989; Dokras et al., 1993; Gardner et al., 2004). However, despite a large number of published studies, there is no consensus about the most accurate method for embryo quality assessment. There are a number of morphological features that display some association with embryo viability, but it is not always clear how these features should be weighted relative to one another. Additionally, the available grading systems rely mostly on visual information obtained by the embryologist and are thus subject to inter-observer (and to some extent intra-observer) variance (Bendus et al., 2006).

Automated image analysis may add objectivity to the process of embryo selection and, consequently, lead to an improvement in the identification of viable embryos for uterine transfer. Several works concerning automated image analysis of cleavage stage human embryos (Day-1 to -3 post-fertilization) have been published (Peder sen et al., 2003; Kaelsson et al., 2004; Morales et al., 2008; Santos Filho et al., 2010a). However, to the best of our knowledge, no attempts at full automated image analysis of human embryos at the blastocyst stage (day 5/6) have been made previously. Designing algorithms that provide a robust analysis of human blastocysts is challenging owing to the subtle nature of morphological differences seen at this stage.

This article presents computerized methods for delineating the boundaries of human blastocyst images, characterizing key morphological features and image classification with the aim of automating blastocyst grading. In the earlier work, we reported on segmentation of the TE and the extraction of its fractal dimension (Santos Filho et al., 2010b). The current article significantly extends on that preliminary work, by presenting the full feature extraction procedure and classification process. We present the first pre-clinical results using our approach and suggest some interesting avenues of future work.

Methods

Image acquisition

The blastocysts obtained following IVF were photographed using the Fer tIMorph system (Image House Medical A/S, Copenhagen). The automatically controlled image recording and storing took ~15 s per sequence. A Nikon optical microscope (×400 magnification and a numerical aperture of 0.55) with Hoffman modulation contrast was used.

Blastocyst grading

An image of an embryo at the blastocyst stage and indication of its structures is shown in Fig. 1a. The blastocyst stage is characterized by the formation of a fluid-filled cavity (the blastocoel) that occupies the centre of the embryo, surrounded by a single layer of cells (the trophodectoderm—TE). Additionally, a small protuberance of cells (the inner cell mass—ICM) that will eventually form the fetus may also be visible. Blastocysts are surrounded by an external glycoprotein membrane, the zona pellucida (ZP), until the time of hatching, when the embryo expands and eventually emerges from the ZP. The morphological grading system most commonly employed at the blastocyst stage is that of Gardner et al. (2004). In this scheme, blastocyst development is ranked from 1 to 6; the ICM quality as A, B or C; and the TE quality as a, b or c.

Blastocyst image segmentation

Image segmentation is the technical term used to describe dividing an image into its meaningful constituent regions. It is one of the first tasks in the process of automatic image analysis. After the image segmentation, features are extracted from the delineated regions and used for identification, measurement or classification. Derived morphological measures of the embryo may be stored in a database to improve knowledge about early embryo development. For instance, based on measurements of ICM, Richter et al. (2001) showed that blastocysts with relatively large and/or slightly oval (roundness index 1.1–1.2) ICMs are more likely to implant than other blastocysts. With a full automatic quantification of blastocyst morphological features, more data of this kind will hopefully become available.

In our approach, ZP inner boundary segmentation is performed by the direct least square fitting of ellipses method proposed by Fitzgibbon et al. (1999). This process is illustrated in Fig. 1a–d. For the segmentation of the ZP outer boundary, an intensity-based threshold does not yield accurate results because the ZP is relatively transparent and does not have a high-contrast outer boundary at the blastocyst stage (earlier embryonic stages are likely to be much easier to assess). Thus, a thresholding based on a method called phase congruency (Kovesi, 1999) is performed to obtain the binary image shown in Fig. 1g.

The variational level set algorithm proposed by Li et al. (2005) is employed for segmenting the TE inner boundary. In this segmentation method, a curve is iteratively evolved outward until it meets the inner boundary of the TE, as illustrated in Fig. 1j and k. However, because of the transparency of the embryo, as well as occasional floating cells in the blastocoel and debris in the neighbourhood, the curve does not always meet the real TE boundary completely, as can be seen in Fig. 1l, in comparison with the boundaries which were manually segmented.

The variational level set algorithm proposed by Li et al. (2005) is employed again for segmentation of the ICM. In this case, the initialization curve was a circle of small radius positioned approximately at the centre of the ICM. The initial curve is then iteratively evolved outwards to the ICM boundary. If the ICM is too far from the embryo centre, the curve is likely to fall outside the ICM and, consequently, the algorithm should be initialized manually by the user.

Feature extraction

After the detection of the TE inner and outer boundaries, a binary image is created. This is converted via a polar coordinate transformation to make the circular ‘band’ of the TE into a horizontal straight strip. The vertical distance between the boundaries is known as ‘TE thickness’. This is calculated and used to construct the final TE thickness signal (Fig. 2c), which is normalized by dividing it by the blastocyst radius length. The procedure reduces the problem of TE feature extraction to the analysis of a one-dimensional (1D) signal, which makes the method faster and allows for the application of a wide range of 1D signal analysis methods to characterize TE thickness.

In order to assess the regularity of the TE thickness signal, the fractal dimension of this signal was calculated. The fractal dimension is a measure of the regularity of a shape. A high-quality TE composed of a large number of cells generates a more irregular and complex TE thickness signal that, consequently, presents a higher fractal dimension. The method of fractal dimension calculation proposed by Higuchi (1988) was used for its simplicity and robustness to signals comprised by a short sequence of points.
Another feature used to characterize the TE is its average thickness, which is the mean value of the TE thickness signal. The TE mean thickness is a useful indicator of TE quality, because as the embryo develops the TE tends to be thinner. The fractal dimension of the TE and mean thickness of TE were also used for characterization not only of TE but also of the blastocyst development stage. The TE mean thickness is correlated with the blastocoel volume/blastocyst volume ratio.

Unlike TE, the ICM region is better characterized by its texture than by its shape. Thus, texture spectrum (He and Wang, 1991) and grey-level run length-based (Galloway, 1975) features were used owing to their efficiency and simplicity. They are used to assess patterns of grey-level arrangements in small areas that compose the global texture of the ICM region. Examples of this image processing are shown in Fig. 3. The texture spectrum-based feature used was the black and white symmetry, which is a measure of the symmetry of the region of interest histogram. The grey-level run length-based feature was the long-run emphasis, which is a measurement of the frequency of occurrence of long runs of a given grey level over the region of interest.

Blastocyst classification

Support vector machine (SVM) is a method that performs classification tasks by determining a separation rule between two sets of feature values. It uses the values around the frontier between the sets. These values are considered the support vectors. In a study presented by Meyer et al. (2003), SVM was compared with 16 classification methods and showed mostly good performances. Thus, we have used SVM as the classifier for our blastocyst grading method.

In a divide-and-conquer fashion, the following classifiers were trained: one classifier to grade development quality, one to grade ICM quality and one to grade TE quality. The features used were those described above and their respective feature spaces can be found in the Supplementary Figures 1–3.

In the blastocyst development feature space (Supplementary Figure 1) a reasonable separation between blastocysts of grade 2 and 4 was observed. Blastocysts of grade 2 tend to present the mean thickness of TE values >0.2 whereas blastocysts of grade 4 tend to present values <0.2.
Regarding the values for TE fractal dimension, despite the lack of clear separation, the blastocysts of grade 2 tend to present values of TE fractal dimension <1.4 (7 out of 9), whereas the blastocysts of grade 4 tend to present values >1.4 (36 out of 51). As the blastocysts of grade 3 occupy a more central position, they mix partially with the others. This makes the classification of grade 3 blastocysts a more challenging task.

The ICM feature space presents visible separation between most of ICM grade B and C, whereas the ones with grade A are more spread out (see Supplementary Figure 2). A similar situation can be observed in the TE feature space (see Supplementary Figure 3). Thus, the classifiers were trained to classify the most frequent and more separated groups present in our data set: development grade 2, 3 and 4; ICM grade B and C; and TE grade b and c.

In order to obtain a full classification of a blastocyst, two images are needed: one with focus on TE (usually on the blastocyst equator) and one image with focus on the ICM. Using a PC with 1.4 GHz processor and 3 GB RAM, the algorithm takes, on average, 4 min to classify a blastocyst.

Establishing a gold standard

In order to evaluate the classifier accuracy, a set of images for embryos of confirmed grade was needed. The images used for this purpose had received the same grade from two independent experts and were therefore considered to be accurate (called ‘gold standard’). While attempting to identify gold standard images, it was noted that the experts agreed on the grading of extent of development for 81% (93 out of 115) of the images. In grading ICM, they agreed for 58% (36 out of 62) of the images. In grading TE, they agreed for 70% (69 out of 98) of the images. This difference in opinion could possibly be reduced if more images from each embryo and from different focal planes were available.

All the algorithms and statistics presented in this paper were implemented or calculated using MATLAB programming language (The Mathworks, Inc., Natick, MA, USA).

Results

Of the 93 images of different blastocysts, for which both experts had agreed on the morphological score for developmental stage, 69 also received an agreed grading of the TE and 36 received concordant scores for the ICM. Disagreements in the morphological scores assigned by the experts were most likely due to imperfect acquisition of images, resulting in suboptimal contrast, or the inclusion of early blastocysts, which often do not present a well defined ICM and TE. After exclusion of embryos with discordant or less frequent scores, the numbers of usable images were reduced to 73 for the training of the ‘developmental stage’ classifier, 54 for training the TE classifier and 20 for training the ICM classifier.

The group of images used for training of classifiers was the same group as used for the test. The leave-one-out method (Kohavi, 1995) was used, as it is suitable for assessing classification accuracy, particularly when the number of available samples is small. It uses a single sample from the training set for test and the remaining ones as training set. This is repeated such that each sample in the training set is used once as a test sample. Using this evaluation method, we applied our grading method to 73 images and the results for assessment of embryo development are shown in the confusion matrix in Table I. A confusion matrix is a table whose rows show actual
classes and the columns show predicted classes done in the classification process.

In tests reported in Santos Filho et al. (2010b) with 25 manually segmented TEs (grade 4 \(n=18\)), grade 3 \(n=5\) and grade 2 \(n=2\), a Pearson correlation coefficient of 0.81 between the fractal dimension and the number of TE cells in the plane of focus was found. This indicates that the fractal dimension is capable of providing a numerical score representative of cell number. The result of this test is shown in Fig. 4.

In a test of full automatic segmentation of the ZP in 44 blastocyst images (27 of grade 4, 8 of grade 3 and 9 of grade 2), we calculated the Dice similarity coefficient (DSC) between automatic segmented ZP and those manually segmented under the supervision of an expert embryologist. DSC, proposed by Dice (Zou et al., 2004), is a measure of spatial overlap between two segmented regions of an image. It ranges from zero, indicating no overlap, to 1 indicating complete overlap. The results obtained are summarized in Table II. These

---

**Table I** Confusion matrices of the classifiers

<table>
<thead>
<tr>
<th>Development</th>
<th>Classified as grade 2</th>
<th>Classified as grade 3</th>
<th>Classified as grade 4</th>
<th>Classifier accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>9 True grade 2</td>
<td>6</td>
<td>3</td>
<td>0</td>
<td>0.67</td>
</tr>
<tr>
<td>13 True grade 3</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>0.46</td>
</tr>
<tr>
<td>51 True grade 4</td>
<td>0</td>
<td>4</td>
<td>47</td>
<td>0.92</td>
</tr>
<tr>
<td>ICM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 True grade B</td>
<td>6</td>
<td></td>
<td></td>
<td>0.67</td>
</tr>
<tr>
<td>11 True grade C</td>
<td>2</td>
<td></td>
<td></td>
<td>0.82</td>
</tr>
<tr>
<td>TE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 True grade b</td>
<td>9</td>
<td></td>
<td></td>
<td>0.53</td>
</tr>
<tr>
<td>37 True grade c</td>
<td>3</td>
<td></td>
<td></td>
<td>0.92</td>
</tr>
</tbody>
</table>

The rows show actual classes and the columns show predicted classes. ICM, inner cell mass; TE, trophectoderm.
results mean that the full automatic ZP segmentation achieved an accuracy level high enough to be used in the estimation of ZP thickness in $\approx 50\%$ of the images tested. Some examples of these results are shown in Fig. 5. For the remaining images, which presented large artefacts or debris, highly accurate segmentation was still achievable but a semi-automatic approach, involving manual initiation of the ellipse fitting algorithm, was necessary.

The development classifier achieved classification accuracy of 0.67 (6/9) in grading embryos defined by embryologists as grade 2; 0.46 (6/13) accuracy in categorizing embryos of grade 3; and 0.92 (47/51) in classifying embryos of grade 4. For ICM classification, as shown in Table I, the accuracy achieved was 0.67 (6/9) for grade B and 0.82 (9/11) for grade C. For TE classification, as shown in Table I, the accuracy achieved was 0.53 (9/17) for grade b and 0.92 (34/37) for grade c.

The distribution of the values of mean TE thickness and mean ZP thickness for a set of 44 blastocyst images of development grade 4, semi-automatically segmented, is shown in Supplementary Figure 4. It is clear from this graph that the values, while tending to be distinct from embryos of other grades, are nonetheless distributed over a range. This emphasizes that embryos within a single grade are subtly different from one another in terms of these morphological features. The variations, although relatively small, are likely to have some prognostic significance and may therefore assist in embryo selection. However, the differences are on a scale that would be extremely difficult to reliably quantify by eye and therefore require digital image analysis.

Discussion

The introduction of automated morphological analyses of embryos, in conjunction with IVF treatment, is an attractive possibility. Such an approach may allow more accurate, non-subjective analyses to take place. It has been shown in the current study, as well as in previous investigations, that there is a significant variation in assignment of embryo grades between different embryologists (Bendus et al., 2006). The use of well-defined algorithms for the evaluation of embryo images would allow morphological scoring to be standardized between laboratories worldwide. In turn, this would permit information to be readily and reliably shared between laboratories, increasing the potential for collaborative research into factors affecting embryo development and assisting in the fine tuning embryo scoring schemes that predict outcomes based on morphology. For these reasons, a number of researchers have sought to investigate the possibility of developing methods for the computer-aided evaluation of embryos (Hnida et al., 2004; Hnida et al., 2005; Beuchat et al., 2008).

Automation of scoring, combined with new incubator systems that include inbuilt microscopes and image capture capabilities, offer a range of exciting possibilities for the future, including more rapid analysis of embryos; evaluation of embryos at additional time points (e.g. during the night); minimization of embryo removal from the carefully controlled environment of the incubator (Lemmen et al., 2008; Wong et al., 2010). Computer-assisted analysis also offers the possibility of much more subtle quantification of individual morphological
features, potentially allowing embryos given identical scores using traditional evaluation schemes to be sub-divided. Culture of human embryos to the blastocyst stage is becoming more common due to advances in physiological media and laboratory techniques. The bottleneck in developing fully automatic blastocyst analysis systems is the image analysis f. This is the focus of the current paper.

The results obtained with the image analysis algorithms reported in this study are encouraging, because they show that a reasonable degree of accuracy in the segmentation of the ZP can be achieved after fully automated analysis of the blastocyst stage. This approach allows the measurement of ZP thickness, which is reported to be an important indicator of embryo quality (Gabrielsen et al., 2000). The use of phase congruency (Kovesi, 1999) images for segmentation of ZP and ICM was found to be very effective, revealing inner structures of the ZP that are almost invisible in the original images and making possible estimation of the ZP outer boundary based on thresholding. As the phase congruency is independent of image contrast, it was possible to find a threshold value that worked well for the 44 well-defined blastocyst images used in the test. The best fitted ellipse method (Fitzgibbon et al., 1999) proved to be appropriate for estimation of the ZP outer boundary based on the blastocyst convex hull vertices. Despite the circular shape of blastocysts, elliptic embryo images may arise mainly because of the orientation of the microscope focal plane. Thus, the fitting of an ellipse makes the algorithm more robust than if a circumference were fitted.

The segmentation of the TE outer boundary, which is the same as the ZP inner boundary, was also successfully achieved. This can be attributed to the robust nature of the ellipse fitting method. The image analysis coped well with difficulties caused by debris and other factors that distort the image, although a uniform background was beneficial for facilitating the thresholding process. The results demonstrate that the method could overcome a well-known limitation of the level set approach, which is that it can be easily disturbed by the presence of distracting objects—in our case the presence of sperm or adherent cumulus cells in the neighbourhood of the embryo.

TE inner boundary and ICM boundary segmentation using level sets worked well but depend on a good initialization. Further work will look at improving on this. In images where the TE or ICM present with weak edges (i.e. boundaries of poor contrast) the segmented boundary can ‘leak’. One solution may be to use a combination of region-based and gradient-based level sets. In the current implementation, a manual initialization is used for the more difficult cases. During the course of this study, no attempt was made to optimize microscope settings or to produce raw images of maximum contrast. It is likely that simple adjustments of the microscope would yield images more amenable to image analysis (i.e. greater contrast and stronger boundaries of individual features). Thus, it is probable that the good results reported here can be improved further, even without any changes to the algorithms.

The analysis of the TE in two dimensions is possible to consider in future work. However, the primary interest is in estimating the number and the regularity of the TE cells, and the analysis of TE in one dimension via the TE thickness signal as used in this work is simpler, faster and more straightforward. There was good correlation between the fractal dimension of the manually segmented TE inner boundary and the number of cells in the TE section. Further work will look at how to extend the classification of TE beyond the three possible groups (a, b and c) in current use by combining estimated cell count and thickness with other measures.

The ICM image texture seems to provide the most useful information for classification of ICM because the ICM region often does not present an image with well-defined boundary. This makes it difficult to extract shape features from the ICM region boundary. The texture features used, based on grey-level run lengths and texture spectrum, seem to be effective ICM texture descriptors. However, a large number of ICM images are needed for a more complete evaluation of their effectiveness.

The results suggested that SVM is an appropriate technique for grading blastocysts. However, a classification tree should also be tested in the future as the appearance of blastocysts changes substantially from one given stage/grade of development to another. In the next steps of this research, classifiers for other grades (TE grade a, ICM grade A, development stage grades 1, 5 and 6) and more information about ZP should be introduced in this proposed method.

**Conclusions**

Automated image analysis may increase the objectivity of embryo morphological evaluation, as well as allow assessment at additional time points when IVF laboratories are not typically staffed. Increased accuracy of measurements may allow embryos to be categorized in a more continuous manner rather than attempting to fit them into arbitrary, discrete grades. The measurement of ZP and TE thickness may help in assessing blastocysts of the same development grade. Also, the quality of TE may be assessed based on its fractal dimension. The main limitation of the proposed algorithms is that, depending on the blastocyst image quality, user intervention may be needed, characterizing a semi-automatic approach. Also, more complete and robust classifiers are necessary. This is likely to be obtained with a larger training set.

In order to determine whether or not this method of embryo assessment truly provides the anticipated clinical benefits, the algorithms will need to be tested against classical human observation in a randomized trial with well-defined clinical endpoints. This is the only reliable way of assessing their true clinical relevance. However, the results obtained using the current set of algorithms are promising, showing that semi-automatic grading of human blastocysts is feasible and suggesting that with more work in this area fully automatic methods may be available in the foreseeable future. Given the growing interest in time-lapse embryo analysis, automated tools for morphometric analysis seem set to have an increasing role in the embryology laboratory.

**Supplementary data**

Supplementary data are available at http://humrep.oxfordjournals.org/

**Authors’ roles**

E.S.F.: paper writer and algorithms developer; J.A.N.: project definition, input to algorithm design, article reviewing; D.W.: concept, assistance with paper writing; M.P.: picture procurement and assessment of blastocyst images; T.G.: assessment of blastocyst images; G.E.: images provider, manual grading and paper reviewers.
Funding

This work was supported by the NIHR Oxford Biomedical Research Centre.

Conflict of interest

None declared.

References


