Logo Recognition with an Incremental Learning method and Consensus for enabling Blockchain Implementations

Josep Lluis De La Rosa, Andres El-Fakdi, Fabio Bacchini and Xesca Amengual Gaya

April 24, 2019

Logo Recognition with an Incremental Learning method and Consensus for enabling Blockchain Implementations

Josep Lluis de la Rosa i Esteva, Andres El-Fakdi, Fabio Bacchini & Xesca Amengual

TECNIO Centre EASY, Vicorob Institute, Universitat de Girona & Università di Sassari

Abstract. A large amount of non-processed visual information is uploaded in social networks every day. Different features can be analyzed from the images such as objects, scenes, sentiments, people's mood, color, etc. In this paper, we propose a novel method to detect, locate and classify logos in images, based on consensus. First, we present a basic logo recognition method. Second, an incremental learning algorithm is proposed to detect logos of any class by just using a synthetic image template, without the need of annotating a training set. Then, a crowdsourced solution (collaborative network) is generated within a VisualAD platform to carry out the consensus between several executions of the incremental learning method. The predictions will be the result of individual predictions from several users that improve the recognition. Finally, the principles enabling its Blockchain implementation are set and considerations on their extension to visual identity

1. Introduction

Logos can be seen as a particular class of object and most of the algorithms used for object detection and localization can be definitely applied to solve this task. Within this context, many different approaches can be found exploiting several kind of features: histograms of edge pixels (Kato, 1992); as union of solid or line-like regions and described by contours and skeletons (Leung and Chen, 2002); as global Zernike moments combined with local features such as local curvatures and distances to centroid (Wei et al., 2009); using global features, like color histograms (Aldershoff and Gevers, 2004), multidimensional receptive field histograms (Phan and Androutsos, 2010), in a variant of the shape context descriptor (Rusinol and Llados, 2010); or in applications based on mobile devices (Kuo and Chen, 2011).

There are some computer vision issues to face to deploy mining of photos in the social web regarding logo detection. Pictures are normally shot without thinking about brands or logos. They will normally appear blurred and occluded which will make its detection a bit more difficult than if we try to detect over controlled visual data. Moreover, it is important that the number of brands and logos to detect in the world is enormous; so we will elaborate a strategy to avoid the need of pre-empting the list of possible classes.

We propose here a new method in the state of the art to detect, locate, and classify logos in images. First, we present our proposal to address the logo recognition problem. Secondly, an incremental learning algorithm is proposed to proceed with a dynamic training set that is originally created by a single synthetic template and run-time filled;

this technique avoids the need of previously building an annotated training set with limited classes. Then, experiments with different settings are carried out and detailed in Section 4. In Section 5, we propose the Consensus strategy by taking advantage of P2P validation architectures initially inspired from the blockchain; it consists of taking into account the predictions obtained by several nodes of the protocol installed in different users from a social network when running our incremental learning method for logo recognition with randomly different training sets. The predictions will be the result of individual predictions from several users that improve the recognition. Finally, the principles for a Blockchain implementation are set in Sections 6 and 7.

2. Logo Recognition

The problem we are addressing consists on classifying images that contain a query logo and indicating where the logo is located. Our logo recognition method is based on the extraction of interest points from the query logo images and the testing images, then, we find matchings between the key points and apply RANSAC to take into account the spatial distribution of those points and discard key-points matchings that are not spatially consistent. The structure of this method is represented in Figure 1.
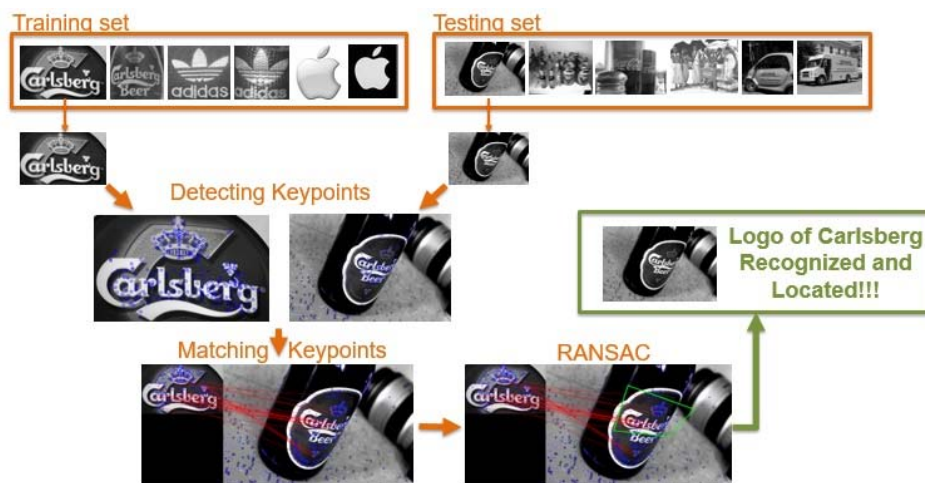

Figure 1. Basic logo recognition method.

We use SURF (Bay et al., 2008) and SIFT (Lowe, 1999) to detect key points and extract the correspondent descriptors. First, SURF descriptors are detected and extracted from the query images; second, the mean number K of key points detected in the images of the same class/logo is computed; third, if K is lower than a threshold, SIFT descriptors are used, instead of SURF, to look for the logo.

This strategy reduces the number of false positives because SURF detects less key points than SIFT and then, less wrong matchings are counted. Also, working with less key points reduces the computation cost. Then, it is worth using SIFT when SURF key points are not enough, but not otherwise.

In order to take spatial information into account, the RANSAC algorithm is used (Fischler and Bolles, 1981).

Our future research will be addressed to improve the proposed logo recognition method by using bag of words and an inverted index to speed up the matching process. In addition, new techniques to incorporate spatial information will be explored. Finally, color information could help to discard false positives.

3. Incremental Learning

Our method is focused on working with images from social networks that contain a high variety of images, and therefore, a high diversity of logo classes. The existing methods need a previously annotated training set that contains a limited number of different logo classes. Then, the use of pre-processed training sets limits that method to a finite number of classes and involves an expensive human work of image selection and annotation.

An incremental learning algorithm will solve the problem by recognizing logo classes with just an initial template image (!). The basis of our proposal is (i) to use an initial logo template with its corresponding wrapping process and (ii) to use the run-time detected logos as new training images. An illustrative scheme is represented in Figure 2.
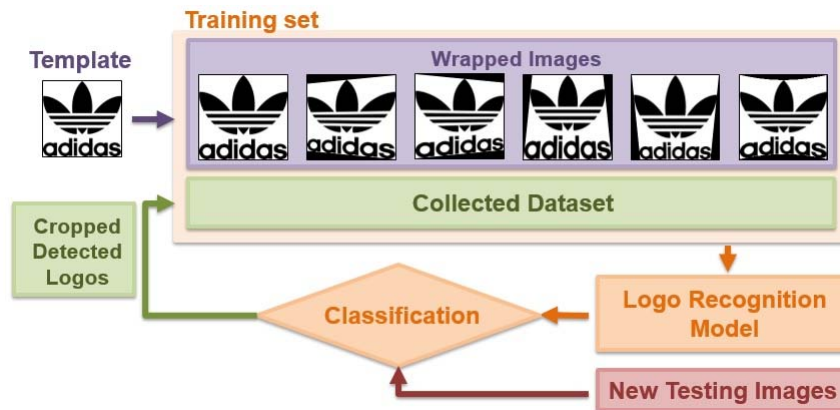

Figure 2. Incremental Learning method.

The first novelty of our method is the use of an initial training set composed by wrapped images from a synthetic image of a logo template. This strategy let users simply take a template of any logo class and build the training set by applying the wrapping transformations. Therefore, any desired logo class can be detected from the testing images without the need of building a training set.

Second, while the method is running, every new logo recognized and detected is cropped from the test image and added to the training set. Then, a dynamic training set is built at run-time, which includes the synthetic wrapped logo templates jointly with the new detected logos 'in the wild'. This provides a richest training dataset with a high variety of types of images and also logos in different environments (perspective, illumination, occlusion, etc.).

In addition, a reinforcement learning method (Sutton and Barto, 1998) is implemented to select the best training images and discard the useless ones. This consists on rewarding the training images that have been very useful on the task of detecting new logos and penalizing the others, then, stablishing the proper thresholds, we select which images are the best to remain in the training set for the next step. The strategy of selecting the best

images, instead of adding every new detected logo, reduces the false positives as well as the executional cost. Moreover, this strategy reduces the probability of corrupting the training dataset by adding wrongly classified logos.

4. Experiments

4.1. Dataset

This method has been designed to be tested over social networks datasets collecting photos that have been taken in different environments and are far from facilitating the logo recognition task. Therefore, we use the public dataset FlickrLogos-32 (Romberg et al., 2011) that contains photos depicting logos. Images are downloaded from Flickr and collect logos of 32 different classes: Adidas, Aldi, Apple, Becks, BMW, Carlsberg, Chimay, Coca-Cola, Corona, DHL, Esso Erdinger, Fedex, Ferrari, Ford, Foster's, Google, Guiness, Heineken, HP, Milka, Nvidia, Paulaner, Pepsi, Ritter Sport, Shell, Singha, Starbucks, Stella Artois, Texaco, Tsingtao and UPS. The whole dataset is split into three disjunct (disjoint?) subsets P1, P2, and P3, each containing images of all 32 classes. A brief summary of the data subsets is shown in Table 1. In addition, we have generated the P1-crop dataset that contains the cropped logos of the images from P1. For the experiments of the logo recognition method we have used P1-crop as a training set and, like (Romberg et al., 2011), P3 for testing.

Subset	Description	Images	Sum
P1	Hand-picked images, single logo, clean background	10 per class	320
P2	Images showing at least a single logo under various views Non-logo images	30 per class 3000	3960
P3	Images showing at least a single logo under various views Non-logo images	30 per class 3000	3960

Table 1. Disjoint subsets of our dataset

For testing the incremental learning algorithm, we build the Template Dataset (TD) that contains the logo template of every logo class. For the sake of comparison, we have selected logo templates of the same 32 classes of the FlickrLogos-32 dataset. Then, we use the images from the FlickrLogos-32 for testing: we divide the FlickrLogos-32 into three subsets I1, I2 and I3. I1 and I2 contains 20 images per class (I1 contains the images from P1 plus 10 images per class from P2) whereas I3=P3. With these three datasets we can run the incremental learning three times, being P3 the last dataset to ease a proper comparison.

4.2. Evaluation Methodology

To evaluate the results of our method we provide the precision and the recall of the whole testing set once all the images have been classified. To have an overall metric to compare the results, we also provide the F score that merge the results of Recall R and Precision P according to the following equation:

$$F = 2 \cdot \frac{P \cdot R}{P + R}$$

4.3. Results

We run several experiments to test our method, split into two parts: first, we evaluate the basic logo recognition method and then, the incremental learning algorithm. Results are shown in Table 2.

Regarding the logo recognition experiments, Table 2 shows the results using only SURF descriptors (LogoRecog_SURF) and combining SURF and SIFT when necessary (LogoRecog_SURF+SIFT). As expected, results are better for SURF+SIFT because using SURF, in a first instance, enables the detection of logos with a lot of key points avoiding false positives and a high computational cost and. Moreover, the use of SIFT in a second instance enables the detection of the logos with less detected SURF key points for which SURF is not enough.

Setting	Recall	Precision	F score
LogoRecog_SURF	0.66	0.84	0.74
LogoRecog_SURF+SIFT	0.72	0.92	0.81
LogoRecog_Template	0.61	0.87	0.72
IncrAddTrainImgs	0.78	0.81	0.80
IncrReplaceTrainImgs	0.77	0.81	0.79

Table 2. Recall and precision results

According to the results of the recognition method, all the experiments of the incremental learning have been executed using SURF+SIFT. Table 2 shows the result of two different settings of the algorithm:

(i) IncrAddTrainImgs: every new detected logo is added to the training set.

(ii) IncrReplaceTrainImgs: according to the reinforcement learning strategy, only the best training images remain in the training set.

For the two strategies we use the P3 subset for the last step, then the Recall and Precision results of the classification are comparable with the other results. In this case, the results are better when adding every new image to the training set but, however this strategy is not feasible when is applied for large datasets, which will be our case, because the training set will become huge and unmanageable. Anyway, we can observe that the difference between the two strategies is very small, so the results with IncrReplaceTrainImgs are still good.

Moreover, we also perform experiments of the basic logo recognition method using the wrapped templates as training images and using SURF+SIFT descriptors (LogoRecog_Template). As expected, the results are worse when using synthetic templates instead of images with logo 'on the wild' because the logos of the testing dataset (P3) are 'on the wild' as well. However, we can observe that the results improve with the incremental learning, then, the initial use of the templates enable the construction of a proper training set containing logos 'on the wild'.

Future work should be addressed to improve the basic logo recognition method, then consequently, the results of the incremental learning would also improve. In addition, the access to larger datasets would allow us a deeper exploration of the reinforcement

learning behavior. Yet the current development is enough to back the necessary modifications for a Blockchain implementation that we see in the next sections.

5. Consensus for Enabling Blockchain Implementations

The incremental learning method for logo recognition lets us detect logos of different classes without the need of building an annotated training. In addition, this method might detect any logo class by just having an initial template; therefore, the number of different logo classes to detect is not limited.

In addition to these advantages, another important benefit comes from the fact that depending on the initial template the predictions will vary which will also result in different subsequent training sets and, consequently, in distinct subsequent predictions. This behavior could be seen as a drawback because the results depend, in part, on the chosen template, however, we see it as the opportunity to compare the different results obtained, so the final predictions, taking into account all the individual results, will be more consistent.

Therefore, we propose to run the Incremental Learning method several times using different templates each time. However, this task has a high computational cost. Then, to tackle this drawback, we also propose to build a network of users to carry out all this computational effort; each user will run the algorithm using a different template and the final results of all users will be filtered by a consensus strategy. The proposed strategy is represented in Figure 3 and consists on the following steps:

1. n users on the internet that gather to validate a logo together $N = \{u_i\}_{i=1}^{n}$
2. They run the Incremental Learning algorithm independently each other
3. Every user u_i selects a template t_i from the Internet. Templates might be all different
4. For the logo class L using the template t_i^L, retrieve a set of images $\mathcal{I}_i^L$ from the test set that contains the logo.
5. For each image I from the test set, we define $\alpha_I = \#\{\, i \mid I \in \mathcal{I}_i^L\}$ as the number of users that have detected the logo class L in the image I. The final set of images that contain the logo class L is defined by
$$\mathcal{I}^L = \{\, I \mid \alpha_I > \varepsilon\,\},$$
where ε is a threshold accordingly set to enhance logo detection recall
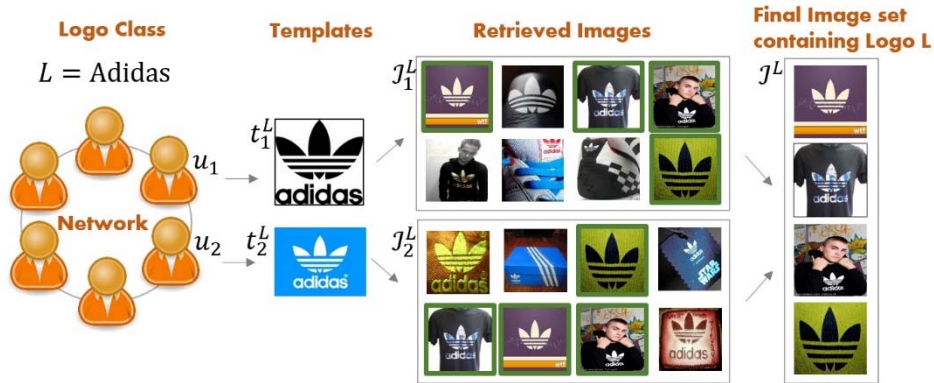
Figure 3. Consensus process

This consensus strategy considers that an image contains a logo only if that logo is detected into the image by a number of users larger than the predefined threshold. Thus, the number of false positives is reduced, so the precision of the results increases. On the other hand, the use of different templates increases the recall because logos that have not been detected using certain templates can be detected by other templates, then, there are more chances to detect all the logos in the images. To achieve a good balance between precision and recall it is important to stablish a proper threshold thr.

It is possible that users cannot contribute to the collaborative network with computational resources. In this case, we provide the possibility to manually filter the results from other users by selecting the correct predictions and discarding the false positives. This collaborative task also aids to improve the results.

5.1 Motivation within the VisualAD platform

When building a collaborative network, it is important to guarantee that users will participate actively, unless the network would not work. We aim to establish a network of users from social networks who will collaborate on the logo recognition task; users will contribute with time and power resources.

Certainly, if users do not receive benefits or rewards for their contributions, the collaborative activity will decrease and the network will fall down. So, we have to define a platform where users can (i) contribute with their resources (computational power or manual annotations) and (ii) receive some benefits for their contributions.

Therefore, we integrate our proposed methodology within the VisualAD platform. This is an application designed to extract visual knowledge from personal images to provide smart marketing analytics. Then, images are assessed according to the information extracted and companies can bid to use them for their interests such as statistics or advertisement strategies. At the same time, the users that own the images used by companies receive benefits from such companies.

So in the context of the VisualAD, users will be more motivated to collaborate with the consensus of the incremental learning method because the retrieved images will be used by companies and the owners of that images will receive benefits and rewards. Then, users are interested on maintaining an active community since their images could be offered to companies and, in that case, they would be rewarded.
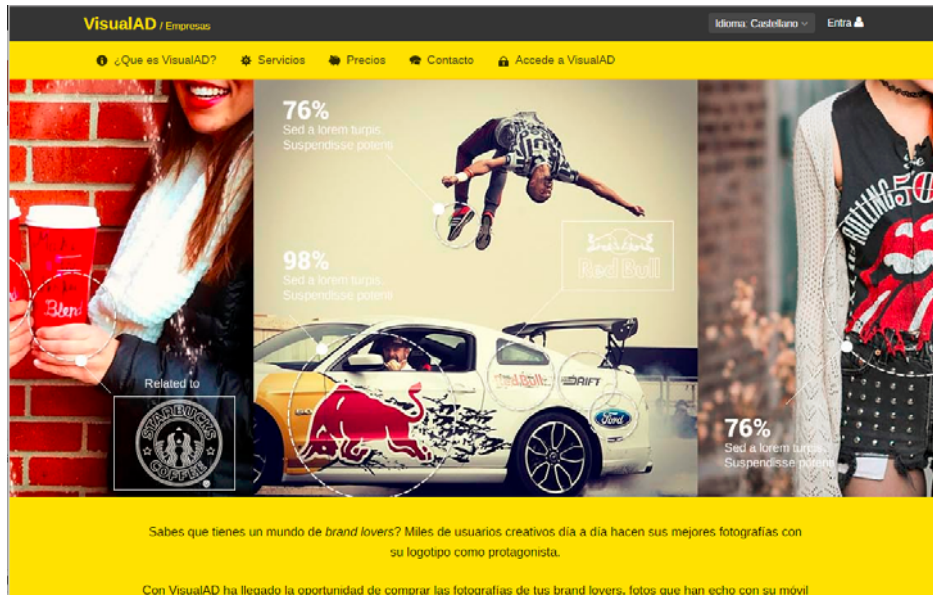
6. A Blockchain Implementation

Blockchain, (Ammous, 2016) (Atzori, 2016) and (DeCovny, 2016) originally the name of the tracking database underlying the virtual currency bitcoin, the term is today used broadly to refer to any distributed electronic ledger using software algorithms to record transactions with reliability and anonymity. This tech. is as well referred to as distributed ledgers (its generic name), cryptocurrencies (the virtual currencies that engendered it), bitcoin (the today prevailing cryptocurrency), and decentralized verification (its key differentiating attribute).

At its heart, blockchain is a self-sustaining, peer-to-peer (P2P) database technology for recording and managing transactions with no central ledger or service involvement. Because blockchain verification is handled through algorithms and consensus among multiple computers, the system is presumed immune to fraud, tampering, or any external control. It is designed to protect against domination of the network by any single computer or group of computers. Participants are relatively anonymous, identified only by pseudonyms, and every transaction can be relied upon.

As well, blockchain and smart contracts convert all computers of the world into 1 universal computer. The distributed ledger technology that started with bitcoin is rapidly becoming a crowdsourced system for all types of verification. The distributed ledger technology that started with bitcoin is rapidly becoming a crowdsourced system for all types of verification. Could it replace notary publics, manual vote recounts, and the way banks manage transactions. Can it be applied to identity as well? A standard Blockchain is indeed a database, a distributed database that relies on a probabilistic consensus mechanism. A Blockchain is also immutable. The consensus-based validation, right at the core operations of the Blockchain, might be the hint to be it applicable to Identity.

7. Proof of Work

Every recognition L validated by α_I users is introduced in the Blockchain and it will enrich the recognition sets of future miners. Every user u_i that detects casts a vote. All votes are recorded in the Blockchain. The Proof of Work consists of the consistent detection of features in the image. Finally, the users get a reward in Visuals, the currency for identity supplied by user as long as they want to be ID validated or paid by entities that require ID for their access

 We are going to apply the "Byzantine fault tolerant" implementation of the Blockchain consensus to our consensus algorithm of Section 5.

1. u_i sends a pic P to the network

2. Candidates $u_j \neq u_i$ mine features of the image $f_j(P)$ (likes *nonces* in the original Blockchain version)

3. u_i easily checks whether $f_j(P)$ matches the properties $f(P)$ and selects users u_k that $u_k \in \{u_j\}$

4. The set $\{u_k\}$ are the ones to participate in the formerly described consensus algorithm for P.

Thus, there is only need to deploy the Proof of Work and Rewards over a network and the "cold start" issue with the first population of images referred to the identities.

8. Conclusions and Future work

We have presented a basic logo recognition method integrated within an incremental learning algorithm that needs only a template of any desired logo class, instead of needing an annotated training set with a limited number of logo classes. It is useful for a consensus strategy that is proposed to provide more consistent predictions as the result of a consensus between all the predictions made by several users using different templates. This consensus is at the core of Blockchain implementations.

 Because of the methodology of the consensus, a collaborative network is generated within the VisualAD platform where users contribute to improve the results whereas receive benefits from interested companies. Part of the rewards are used to pay the costs of the consensus algorithm implemented with Blockchain technology as fees.

 Thus the only need to deploy the Proof of Work and Rewards over a network and the "cold start" issue with the first population of images referred to the logos.

References

F. Aldershoff and T. Gevers, "Visual tracking and localisation of billboards in streamed soccer matches," in *Proceedings of SPIE Storage and Retrieval Methods and Applications for Multimedia*, (pp. 408-416) 2004.

S. H. Ammous, "Blockchain Technology: What is it good for?." Available at SSRN 2832751 (2016).

M. Atzori, "Blockchain-Based Architectures for the Internet of Things: A Survey." Available at SSRN (2016).

H. Bay, A. Ess, T. Tuytelaars and L. Van Gool. "Speeded-up robust features (SURF)," *Computer vision and image understanding*, 110(3), 346-359, 2008.

S. DeCovny, "A Ride around the Blockchain." CFA Institute Magazine 27.3 (2016): 28-30.

M. A. Fischler and R. C. Bolles. "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, 24(6):381–395, 1981.

T. Kato, "Database architecture for content-based image retrieval," in *Proceedings of SPIE Image Storage and Retrieval Systems*, 1992.

S. Kuo and Z. Chen, "Robust Logo Recognition for Mobile Phone Applications", *Journal of Information Science and Engineering*, vol. 27, pp. 545-559, 2011.

W. H. Leung and T. Chen, "Trademark retrieval using contour-skeleton stroke classification," in *Proceedings of IEEE International Conference on Multimedia and Expo (ICME)*, 2002.

D. G. Lowe. "Object recognition from local scale-invariant features," In *Proceedings of the seventh IEEE international conference on Computer vision*, Vol. 2, pp. 1150-1157, 1999.

R. Phan and D. Androutsos, "Content-based retrieval of logo and trademarks in unconstrained color image databases using color edge gradient co-occurrence histograms," *Computer Vision and Image Understanding (CVIU)*, vol. 114, no. 66–84, 2010.

S. Romberg, L. G. Pueyo, R. Lienhart and R. Van Zwol. "Scalable logo recognition in real-world images," in *Proceedings of the 1st ACM International Conference on Multimedia Retrieval* (p. 25), 2011.

M. Rusinol and J. Llados, "Efficient logo retrieval through hashing shape context descriptors," in *Proceedings of International Workshop on Document Analysis Systems*, 2010.

R. S. Sutton and A. G. Barto. "Reinforcement learning: An introduction," *MIT press*, 1998.

C.-H. Wei, Y. Li, W.-Y. Chau, and C.-T. Li, "Trademark image retrieval using synthetic features for describing global shape and interior structure," *Pattern Recognition*, vol. 42, no. 3, pp. 386–394, 2009.