Stock Prediction by Searching for Similarities in Candlestick Charts

CHIH-FONG TSAI and ZEN-YU QUAN, National Central University

The aim of stock prediction is to effectively predict future stock market trends (or stock prices), which can lead to increased profit. One major stock analysis method is the use of candlestick charts. However, candlestick chart analysis has usually been based on the utilization of numerical formulas. There has been no work taking advantage of an image processing technique to directly analyze the visual content of the candlestick charts for stock prediction. Therefore, in this study we apply the concept of image retrieval to extract seven different wavelet-based texture features from candlestick charts. Then, similar historical candlestick charts are retrieved based on different texture features related to the query chart, and the "future" stock movements of the retrieved charts are used for stock prediction. To assess the applicability of this approach to stock prediction, two datasets are used, containing 5-year and 10-year training and testing sets, collected from the Dow Jones Industrial Average Index (INDU) for the period between 1990 and 2009. Moreover, two datasets (2010 and 2011) are used to further validate the proposed approach. The experimental results show that visual content extraction and similarity matching of candlestick charts is a new and useful analytical method for stock prediction. More specifically, we found that the extracted feature vectors of 30, 90, and 120, the number of textual features extracted from the candlestick charts in the BMP format, are more suitable for predicting stock movements, while the 90 feature vector offers the best performance for predicting short- and medium-term stock movements. That is, using the 90 feature vector provides the lowest MAPE (3.031%) and Theil's U (1.988%) rates in the twenty-year dataset, and the best MAPE (2.625%, 2.945%) and Theil's U (1.622%, 1.972%) rates in the two validation datasets (2010 and 2011).

9

Categories and Subject Descriptors: H.2.8 [**Database Management**]: Database Applications—*Data mining*; H.4.2 [**Information Systems Applications**]: Types of Systems—*Decision support*

General Terms: Design, Algorithms, Performance

Additional Key Words and Phrases: Stock prediction, data mining, candlestick chart

ACM Reference Format:
Tsai, C.-F. and Quan, Z.-Y. 2014. Stock prediction by searching for similarities in candlestick charts. *ACM Trans. Manage. Inf. Syst.* 5, 2, Article 9 (July 2014), 21 pages.
DOI:http://dx.doi.org/10.1145/2591672

1. INTRODUCTION

Investment has become an integral part of our daily lives, and many people have become accustomed to using their personal assets to generate more profit. In general, conservative investors place their excess cash in banks, funds, or the stock market, and expect to obtain returns on interest and dividends. In contrast, active investors prefer to use their funds to buy/sell stocks, funds, futures, options, and so on, expecting to get higher returns. As a whole, since stock investment can produce high profitability (although there is a risk), it is one of the most popular investment activities. Although stock investment can bring high returns for investors, the stock market is

This research is partially supported by the National Science Council of Taiwan (NSC100-2628-H-008-001-MY3).
Authors' address: C.-F. Tsai and Z.-Y. Quan, Department of Information Management, National Central University, R.O.C; email: cftsai@mgt.ncu.edu.tw and 984403005@cc.ncu.edu.tw.

DOI:http://dx.doi.org/10.1145/2591672

affected by macroeconomic and international factors, stock market volatility, as well as governmental policies that combine to make it very difficult to achieve an assured level of accurate stock prediction [Araújo 2010; Park and Shin 2013].

The main idea behind stock prediction is to analyze the available information from the stock market in order to effectively predict future trends (stock prices), which can increase profit. Zarb and Kerekes [1970] argued that the main strategy of investment analysis should be based on the stock profile and timing mechanisms offered by fundamental analysis and technical analysis, respectively, which can effectively enhance the return on investment performance. The first approach is aimed at analyzing the general economic situation, current industrial trends, and other long-term factors. On the other hand, technical analysis focuses on the short-term stock price (also called momentum) in the stock market. The latter method [Friesen et al. 2009; Zhu and Zhou 2009] uses some technical indicators and pattern recognition/analysis such as Dow theory and the Elliott line to derive rules for investment strategies, and to predict future stock prices. This approach is the most widely used for stock prediction.

The efficient market hypothesis (EMH) [Fama 1970] maintains that market prices fully reflect all available information, and it is not possible to consistently outperform the overall stock market by stock picking and market timing alone. The fluctuation in stock prices can be represented as a random phenomenon. This hypothesis negates the value of technical analysis. However, consistent fluctuations have been noted in many studies, such as the weekend effect [Doyle and Chen 2009], January effect [Kang 2010; Moller and Zilca 2008], calendar anomalies [Ariss et al. 2011], and week-of-the-year effect [Levy and Yagil 2012]. Brock et al. [1992] used a bootstrapping method to verify that technical analysis can be used to analyze the movement of stock prices, while Hsu and Kuan [2005] used White's reality check and Hansen's SPA test methods [Hansen and Lunde 2005] to show that technical analysis is useful for stock prediction.

Traditional technical analysis is sometimes based on statistical methods, such as auto-regressive (AR) and auto-regressive conditional heteroskedasticity (ARCH) techniques [Engle 1982]. Recently, heuristic and intelligent techniques, such as the genetic algorithm (GA) [Araújo and Ferreira 2013], artificial neural network (ANN) [Hsieh et al. 2011; Liu and Wang 2012; O'Connor and Madden 2006; Qi and Zhang 2008], and support vector machine (SVM) [Luo and Chen 2013; Sapankevych and Sankar 2009] methods, have been applied to predict stock prices. However, in most large-scale statistical finance studies, data are examined in a completely deductive manner, with findings complemented by testing a hypothesis that would be difficult to derive or formulate using purely statistical or intelligent methods. In general, a candlestick chart contains the trading time, volume, and price, and shows the volatility of a stock price within a specific period. In other words, a candlestick chart can fully represent the stock's movement. Hence, candlestick chart analysis has become an important research tool in stock prediction.

In the literature, candlestick chart analysis is widely used for stock price prediction [Caginalp and Laurent 1998; Fiess and MacDonald 2002; Lee and Jo 1999; Lee et al. 2006; Morris 2006]. However, there have also been some studies focusing on candlestick chart analysis that indicate that this approach is not profitable for predicting stock performance [Fock et al. 2005; Marshall et al. 2006, 2008]. In these studies, candlestick chart analysis is usually based on some numerical formulation. Image processing techniques have so far not been used to directly analyze the visual content of candlestick charts for stock prediction. A candlestick chart can be regarded as a binary image, which contains a lot of visual information. In addition, this makes it easy to visualize stock trends and price movements without any calculations. We can combine the traditional raw data (numerical) and visual information (visual vector) to help investors make decisions. We can automatically extract some visual features utilizing image processing algorithms, for further analysis. Note that these visual features are not necessarily interpreted by the human eye; similarity measures are employed to distinguish between different charts. In other words, a user (investor) provides a query chart (e.g., today's candlestick chart) and searches for similar historical charts. This helps them to refer to past stock movements as shown in similar historical charts from which to predict future stock prices. This process is similar to content-based image retrieval (CBIR) [Rui et al. 1999; Smeulders et al. 2000]. In short, the aim of this study is to examine the suitability and applicability of the CBIR technique for stock prediction using candlestick charts.

The rest of this article is organized as follows. Section 2 briefly reviews the literature on candlestick chart analysis. In Section 3, we introduce a new candlestick chart analysis scheme based on the concept of using CBIR to extract related visual features from candlestick charts and to search for similar charts for stock prediction. Sections 4 and 5 present the experimental results and conclusions, respectively.

2. LITERATURE REVIEW

2.1. Stock Price Prediction

Investors usually endeavor to predict stock prices so as to obtain the highest returns in the long term. Technical analysis is a mainstream stock price analysis method. Professional analysts usually consider historical incidents or patterns as replaying again and again [Edwards et al. 2012; Murphy 1999], which will be reflected in stock market movement. For example, the closing prices, trading volume, and trading time are important factors in predicting future stock prices. In general, technical analysis can be further divided into quantitative analysis and graphical analysis. Quantitative analysis uses technical indicators to predict market trends, such as the moving average (MA), stochastic oscillator (KD), relative strength indicator (RSI), and moving average convergence/divergence (MACD), while graphical analysis uses charts for stock prediction (Section 2.2).

In the past, stock price prediction models have been developed based on statistical or regression analysis tools. Auto-regressive (AR), Auto-regressive moving average (ARMA), auto-regressive integrated moving average (ARIMA), auto-regressive conditional heteroskedasticity (ARCH), and generalized ARCH (GARCH) [Bollerslev 1986; Engle 1982; Hung 2009] are nonlinear models used to predict the values and trends of the stock market. More recently, data mining and artificial intelligence techniques have been widely used to develop prediction models, such as GA [Araújo and Ferreira 2013], ANN [Hsieh et al. 2011; Liu and Wang 2012; O'Connor and Madden 2006; Qi and Zhang 2008], SVM [Luo and Chen 2013; Sapankevych and Sankar 2009], and fuzzy ANN (FNN) [Liu et al. 2012; Oh et al. 2006]. In addition, hybrid-based prediction models that combine some of these methods with self-organizing maps (SOM), hidden Markov models (HMM), support vector regression (SVR), particle swarm optimization (PSO), and simulated annealing (SA), have also been constructed [Asadi et al. 2012; Briza and Naval 2011; Cheng et al. 2010; Hsu 2011; Kwon and Moon 2007; Wang et al. 2010; Zarandi et al. 2013]. Tsai and Hsiao [2010] used different feature selection methods to identify more representative variables for better prediction. Moreover, some researchers extracted textual information from financial reports which they used to predict stock price movement [Chan and Franklin 2011; Hagenau et al. 2013; Lin et al. 2011; Schumaker and Chen 2009].

However, it is often necessary when developing complex mathematical models to decide on many variables as inputs to the models, although there is no general agreement about what the most important and representative variables are. Furthermore, it is a massive challenge for users to consider the interplay between these technical

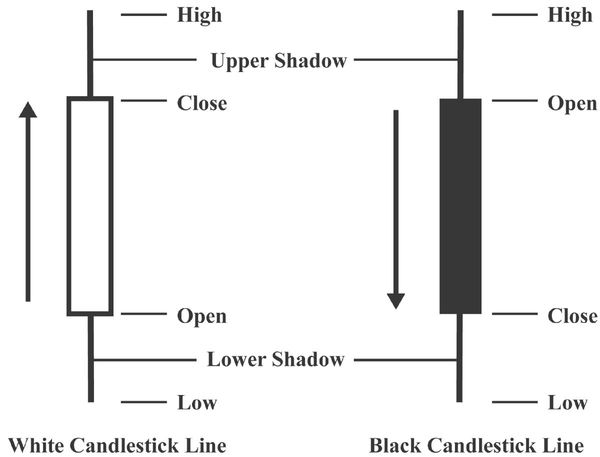


Fig. 1. Candlestick lines representing one trading day.

indicators. This means that some financial expertise is needed in order to interpret the output of these prediction models. In other words, it is very hard for ordinary investors to use technical indicators for stock investment.

2.2. Candlestick Chart Analysis

In addition to quantitative analysis, graphical analysis can be used to predict stock prices. This approach can be further divided into candlestick chart analysis [Caginalp and Laurent 1998; Fiess and MacDonald 2002; Lee and Jo 1999; Lee et al. 2006; Morris 2006], and chart pattern analysis [Chang and Osler 1999; Leigh et al. 2002].

A candlestick chart is one way to look at different eras, which are represented by the amount and types of trading volatility and can be used to explain the relationship between the trading price and the timing. In this type of analysis, it is assumed that the candlestick line or candlestick pattern faithfully reflects the facts (stock price movements) of the stock market. Investors adjust their trading strategies according to the content and trends conveyed by the information. In chart pattern analysis, however, useful patterns that can be used to predict future trends are searched for and identified in the candlestick chart(s) (patterns such as Towers, Wedges, Rounding, etc.). In general, the results of chart pattern studies vary depending on the patterns, markets, and the sample periods tested. The results suggest that some chart patterns might be profitable in stock markets and foreign exchange markets.

Each line in a candlestick chart contains a candle body, and upper and lower shadows. Figure 1 shows examples of white and black candlestick lines, in which each line represents one trading day. The trading time unit is set to be one day. The difference between the opening and closing prices is used to build the candle body on the chart. The entity is white if the opening price is less than the closing price, that is to say, the trend shows a bull market. In contrast, if it is a bear market, the color is black. The upper and lower shadows above and below the candle's body respectively represent the trading range within a specific time period. The investor can interpret the day-to-day sentiment simply by looking at changes in the body color and the length of the shadow on the chart. However, this task is very difficult for general investors.

Morris [2006] found that candlestick charts are suited for short-term investment prediction, and that the most efficient time period for prediction is 10 days. Caginalp and Laurent [1998] treated the S&P 500 as their subject and collected candlestick charts for the period between 1992 and 1996. They then established eight three-day reversal patterns for hypothesis testing. Their results showed that the candlestick charts displayed descriptive and (reversal) trend predictive powers, and that those patterns could provide important information to investors. Lee and Jo [1999] developed an expert system using candlestick charts to predict stock prices and to decide on the best time to buy/sell stocks. Their approach divided the changes in the stock market into falling, rising, neutral, trend-continuation and trend-reversal trends, and they designed 21 rules based on these patterns. Their experimental results showed the average hit rate of these rules to be 72%. However, their expert system required an expert knowledge base that needed to be collected and maintained by knowledgeable engineers. Fiess and MacDonald [2002] showed that the high, low, and closing prices are useful information for forecasting the volatility as well as the level of future exchange rates. Lee et al. [2006] introduced a fuzzy time series data representation method based on the candlestick theory for assisting in financial prediction. In their method, the investment expertise can be stored in the knowledge base, and the fuzzy candlestick patterns can also be identified automatically from a large amount of financial trading data.

On the other hand, there have been some studies that have found that candlestick chart analysis is not useful to assist investors to establish or adjust decisions. For example, Fock et al. [2005] used FDAX and FGBL data for the years ranging from 2002 to 2003 to verify whether candlestick chart analysis could be used to obtain super-normal profits by using t-testing. They also added the MA, RSI, and MACD indicators to increase the accuracy rate. However, their approach showed only limited improvement. Therefore, they concluded that candlestick chart analysis cannot be used to obtain super-normal profit. Marshall et al. [2006] used 35 stocks based on the Dow Jones Industrial Average Index (INDU) to calculate an average indicator for the period from 1 January 1992 to 31 December 2002, and used an extension of the bootstrap methodology to generate random opening, high, low, and closing prices. Their results revealed candlestick chart analysis to have no statistical significance. They concluded that candlestick chart analysis could not correctly reflect the status of the stock market. Similarly, Marshall et al. [2008] reused the same method with the top 100 stocks from historical Tokyo Stock Exchange (TSE) data from 1975 to 2004. They verified that candlestick chart analysis was not helpful for the majority of the stocks for any of the sub-periods whether in a bull or bear market.

However, candlestick chart analysis, as in these studies, cannot avoid the use of interpretable results made by human experts. Such results are subjective and may differ because of the preferences, background, experience, and so on, of these experts. For instance, the analysis results could differ from one time to another without taking into account any fundamental or technical factors. Moreover, the trends and patterns of candlestick charts must be interpreted by human judgment, which involves labor costs and increases the possibility of misinterpretation. On the other hand, many of these candlestick chart analysis studies are based on using numerical formulas to convert the charts into numbers. For example, the relative strength index $(\mathrm{RSI}) = 100 - \left[\frac{100}{1-RS}\right]$, where RS is (average of n periods closing up) / (average of n periods closing down).

Since it is difficult to objectively analyze the trends and patterns shown in the various candlestick charts, we focus here on image processing as a method of analysis. A candlestick chart is a black and white image containing some visual features, such as spectral based texture and shape, and it can represent the trend and volatility of stock prices. These features act as image descriptors and can be automatically extracted by specific related algorithms to describe their visual content. More significantly, the visual features are signals, from the viewpoint of computer vision, that can provide an

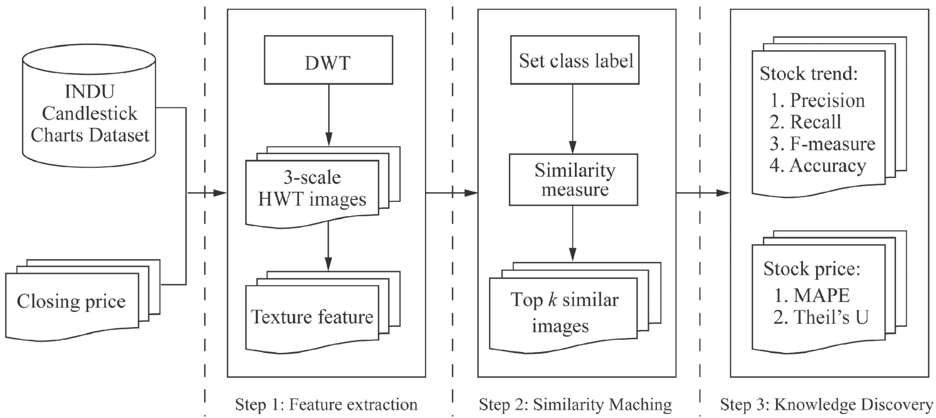


Fig. 2. The process of candlestick chart mining.

objective message for stock price analysts without requiring the experience of financial experts.

It is assumed that the different candlestick charts represented by these descriptors or features can be further analyzed by the similarity matching process for later stock prediction. This turns into a domain problem of content-based image retrieval (CBIR), where a user can provide a query image from which to search for similar images in the image database [Chianga and Cheng 2009; Graña and Veganzones 2012; Lee et al. 2009; Yue et al. 2011]. Therefore, for stock prediction using candlestick charts, investors can search for historical charts similar to the query one, such as today's chart. Then the most similar chart is retrieved and the historical stock movement associated with it can be referred to as indicative of the future stock market trend.

3. RESEARCH METHODOLOGY

Related works of stock prediction use different trading data (stock data in their original form) and some technical indicators to serve as the input values. In other words, these technical indicators are treated like valuable raw materials, used by the prediction models (such as statistical and artificial intelligence methods) for stock prediction. As a result, the hidden knowledge (future stock price movement) from the transaction data can be found and consequently the value of technical indicators can be demonstrated.

In this study, we applied the concept of CBIR to automatically extract texture features from candlestick charts and to retrieve similar historic charts for purposes of stock price prediction. Moreover, texture features can be seen as another type of technical indicator and as valuable resources representative of the trend and volatility of stock prices. Specifically, our candlestick chart analysis method (or candlestick chart mining) is composed of the steps of feature extraction, similarity matching, and knowledge discovery. Figure 2 shows the three step candlestick chart mining process. The steps are described in detail in the following.

3.1. Texture Feature Extraction

3.1.1. Texture Analysis. Texture is an important image feature used to describe the object content and the scene of the images. Moreover, it can express the content structure and the relationship between objects and the surrounding environment. Therefore, texture analysis always plays an important role in computer vision and pattern recognition.

Most traditional statistical approaches to texture analysis, such as first order or second order statistics of textures, Markov random field, local linear transforms, and texture spectrum techniques [Chen and Pavlidis 1983; He and Wang 1990; Krishnamachari and Chellappa 1997], have been restricted to the analysis of spatial interactions over relatively small neighborhoods on a single scale.

Statistical measurement methods [Chen et al. 2011], such as the block difference of inverse probabilities (BDIP), block variation of local correlation coefficients (BVLC), and gray-level difference matrix (GLDM), have already been widely applied in the literature. However, these features cannot faithfully reflect the stock price volatility of every trading day in candlestick chart analysis. Therefore, the goal of this study is to extract the texture features suitable for candlestick chart mining, and to apply the CBIR method to predict stock price movement.

Computer vision research has suggested that texture feature representation must preserve both local and global information. These findings have served as the basis for several approaches to texture analysis using Gabor filters, but much more computational effort is required to extract the texture features and these transformations are usually not reversible, which limits their applicability for texture synthesis. The wavelet transform (WT) method can be used to determine the frequency bands carrying the most information about the texture by decomposing images into multiple frequency bands and computing the band energies. This can be viewed as dimensionality reduction or removing irrelevant data prior to the feature extraction process. Thus, WT offers the ability to achieve robust feature extraction from images. Moreover, it also provides a precise and unified framework for the analysis and characterization of a signal, which could avoid most of the problems in texture analysis [Lilly and Olhede 2010; Quellec et al. 2010].

3.1.2. Discrete Wavelet Transform (DWT). The discrete wavelet transform (DWT) is a popular image processing algorithm that divides the input signal into high-frequency and low-frequency components. The result of decomposition is an alternative representation of image data, but the original information has not been changed [Andreopoulos 2009; Hsia et al. 2009; Wink and Roerdink 2010; Zheng and Huang 2013].

Since an image is comprised of two-dimensional (2-D) data, performing DWT involves two stages: horizontal division and vertical division [Andreopoulos 2009; Hsia et al. 2009]. During DWT, we call the first level DWT image, the subband LL_1, *i.e.*, the coarse overall shape on the upper-left side, which is the low frequency component, and contains most of the energy in the image. LH_1, HL_1, and HH_1 contain the higher frequency detailed information, *i.e.*, the finest scale detailed wavelet coefficients on the bottom-left, upper-right and bottom-right, respectively. The DWT can be applied to obtain the next coarser scale by further decomposing the subband LL_1 into the subbands LL_2, LH_2, HL_2, and HH_2 (each one fourth the size of the subband LL_1). Similarly, if the process is repeated p times, we can get LL_p, LH_p, HL_p, and HH_p, where LL_p contains the information most similar to the original image.

The Haar wavelet transform (HWT) is a well-known transfer function, and is described in the following [Hsia et al. 2013; Zheng and Huang 2013]. An image S with a size of $W \times H$ pixels is decomposed into four respective subbands represented by the following equations:

$$LL_1(i,j) = \frac{1}{4}\sum_{m=0}^{1}\sum_{n=0}^{1} S(2i+m, 2j+n), \tag{1}$$

$$LH_1(i,j) = \frac{1}{4}\sum_{m=0}^{1} S(2i+m, 2j) - \sum_{m=0}^{1} S(2i+m, 2j+1), \tag{2}$$

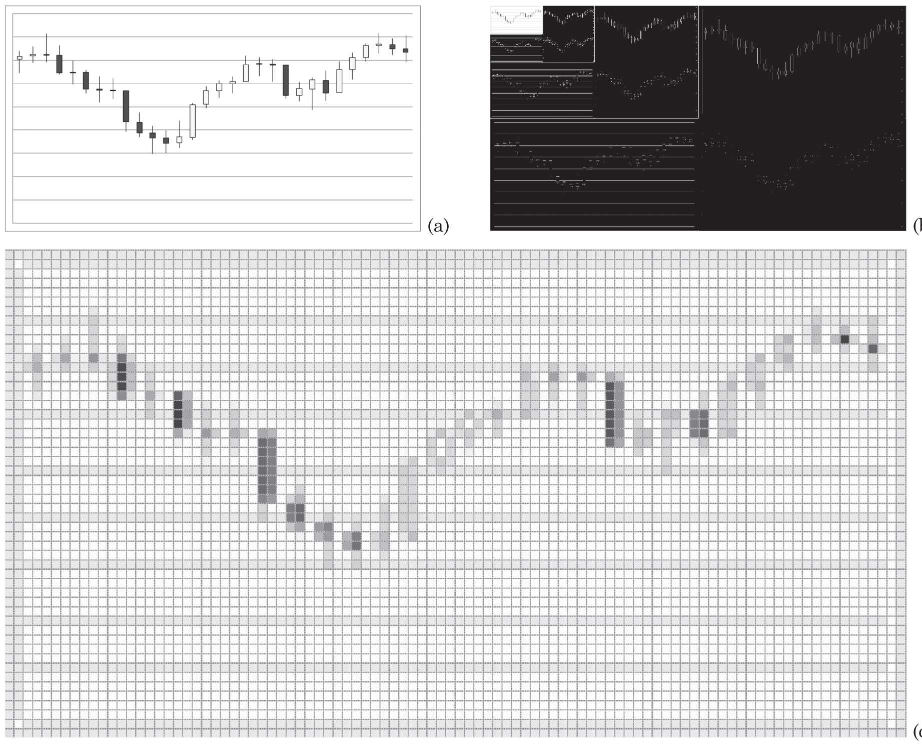


Fig. 3. (a) A candlestick chart; (b) three-scale HWT results; (c) amplified image of the LL_3 section.

$$HL_1(i,j) = \frac{1}{4}\sum\nolimits_{n=0}^{1} S(2i, 2j+n) - \sum\nolimits_{n=0}^{1} S(2i+1, 2j+n), \tag{3}$$

$$HH_1(i,j) = \frac{1}{4}\left[S(2i,2j) + S(2i+1,2j+1) - S(2i+1,2j) - S(2i,2j+1)\right], \tag{4}$$

where $S(i, j)$ is the pixel value on coordinate (i, j) of S if $1 \leq i \leq W$ and $1 \leq j \leq H$; and $LL_1(i, j)$, $LH_1(i, j)$, $HL_1(i, j)$, and $HH_1(i, j)$ are the coefficients on coordinates (i, j) of subbands LL_1, LH_1, HL_1, and HH_1, respectively. Similarly, LL_p, LH_p, HL_p, and HH_p can be computed in the same way.

The spatial resolution increases with the frequency, and the spatial-frequency resolution becomes narrower as the frequency decreases. Hence, the high frequency part of the image is more subtle, such as at the edge of an image; and the low-frequency part is smooth. Sharp edges, which are well localized spatially and have significantly higher frequency content, can be represented more compactly by WT than the other transforms.

3.1.3. Texture Feature Representation. According to the definition of DWT, since most of the information in an image is kept in the p-th low subband (LL_p), we can transfer each candlestick chart into a three-scale HWT, and then extract related features from the LL_3 subband. Let us look at the example in Figure 3(a) in which there are T consecutive trading days in a candlestick chart. Note that every candlestick chart is 768×416 (319,488) pixels in size, and the coordinate axes of the top-left and the bottom-right part are (1, 1) and (768, 416), respectively. For the example of Figure 3(a), the trading data for the first trading day are transferred into the left candlestick line, starting from (26, 83) to (26, 124), with the length of this line being 41 (=124–83) pixels. Figure 3(b) shows the results after the three-scale HWT. Although the LL_3 subband shows the coarse outlines of the candlestick chart, this can fully represent the information in the original. Figure 3(c) shows an amplified image of the LL_3 section.

More significantly, each pixel value in Figure 3(a) is either black (0) or white (255), since it is a black-white image. After HWT processing, the value of every pixel ranges from 0 to 255, in which the value is composited with the length of the upper and lower shadows and the candle body. Thus the different pixel values reflect the ups, downs, and volatility of the stock price. Therefore, LL_3 not only represents spectrum domain data, but also faithfully reproduces textural data.

As a result, we can obtain T pieces of main information from every chart corresponding to the T trading days, each one constituted by an upper shadow, a lower shadow, and a candle body. However, in a candlestick chart the pixel values for this information are only black or white. After performing the three-scale HWT, we find that the pixel values in the LL_3 section are abundant, because we condense the main information. Therefore, we call these features texture-based features (TBFs). In addition, the length of the shadows represents the volatility of the up and down movement of the stock prices, that is to say, a longer shadow means that the volatility of the stock price is more obvious. In contrast, a shorter shadow means that the volatility of the stock price is minor. Based on these observations, we can further extract the status of the rising and falling of stock prices on the relevant dates, while recording the highest and lowest positions. Consequently, we call these features, location-based features (LBFs).

In this study, the extracted feature vectors consist of TBFs and LBFs. For simplicity, we set i_x as the location of the x-th main-information ($x = 1, 2, \ldots,$ and T), while $j_{x,1}$ to $j_{x,y}$ indicate the length of the shadow line, and $k_{x,1}$ to $k_{x,y}$ indicate the length of the candle body. In addition, we extract four pixel values from the TBFs that are the average pixel values of the whole shadow line (AS_x) by

$$AS_x = \sum\nolimits_{r=1}^{y} LL_3(i_x, j_{x,r})/y; \tag{5}$$

the pixel values of the highest location of upper shadow (US_x) by

$$US_x = LL_3(i_x, j_{x,y}); \tag{6}$$

the lowest location of the lower shadow (LS_x)

$$LS_x = LL_3(i_x, j_{x,1}); \tag{7}$$

and the middle point of the candle body (CBS_x)

$$CBS_x = LL_3(i_x, \frac{\lfloor k_{x,y} - k_{x,1} \rfloor}{2}). \tag{8}$$

In addition, the positions of the upper shadow (ULP_x) and the lower shadow (LLP_x) are extracted to represent the LBFs. Consequently, there are six values describing the information for a trading day, which results in $6\times T$ features describing each candlestick chart.

3.2. Similarity Measure for Chart Retrieval

Once the texture features are extracted, the next step is to retrieve historical charts that are similar to a given query chart, which is done by similarity measurement. In CBIR, each image is usually represented by an N-dimensional feature vector, which can be regarded as a point in N-dimensional space. The retrieval of images similar to a query image depends on the similarity between the image features stored in the image database and the query image features. The level of similarity between any two images depends on the distance in the feature space between the feature points represented by the descriptors. Therefore, the shorter the distance between the two points, the more similar the corresponding images. A number of well-known similarity metrics, such as the Euclidean distance, cosine distance, and χ^2 statistics have been used in the literature.

In this study, the Euclidean distance (ED) is used both because it is easy to implement and it can be regarded as the baseline similarity metric, as defined in Equation (9). In this expression, images X and Y are represented by N-dimensional feature vectors, in which X_n and Y_n are the n-th ($n = 1, 2, \ldots,$ and N) elements of the feature vectors

$$ED = d(X, Y) = \sqrt{\sum\nolimits_{n=1}^{N} (X_n - Y_n)^2}. \tag{9}$$

For candlestick chart retrieval, choose the candidate chart having the greatest similarity (the shortest distance between the query chart and the chart stored in the database) as the retrieval result for evaluation. Note that there are some other possible choices when collecting the retrieval results, such as the top 15 most similar candlestick charts that can be simultaneously examined. The class label or prediction output (stock rises or falls) that receives the largest number of votes is selected as indicating the future trend for that stock. However, here, we select the most similar chart, using it to directly assess and justify the discriminatory power of these feature descriptors.

3.3. Retrieval Evaluation

The final step is to evaluate the retrieval results in terms of knowledge discovery for stock prediction. Several methods have been utilized in past image retrieval studies to evaluate the retrieval performance, with the four most widely used measurements being accuracy, precision, recall, and F-measure.

Before evaluating the retrieval results, each feature vector in the image database should be followed by a specific class label. For the first experiment, we simply define two class labels, which indicate the stocks' ups and downs. However, since each chart contains T pieces of main-information (T trading days), the retrieval results could be different if different dates are used as indicators to predict the rising and falling of stock prices as the class label.

The scenario for performing stock prediction is as follows. Suppose today's chart is used as the query to search for the most similar historical chart $C_{m1\sim m30}$ in the database (where m represents a specific trading day); then the next t-days stock movement for $C_{m1\sim m30}$, *e.g.*, $m31$, is used as the next one day stock movement for $C_{m1\sim m30}$ and is used as the result for predicting tomorrow's stock price. Specifically, the trading days are Monday to Friday in every week; Saturday, Sunday, and official holidays are not trading days since the stock market is closed. Therefore, f is a future trading day when transactions have not been suspended because of holidays, and f is set to 1, 7, 15, 30, and 90, respectively, for comparison.

Moreover, sometimes the forecast trend of the rising and falling of stock cannot meet the investors' needs in reality. They may care more about the actual closing price instead of its rising or falling (a binary classification). Therefore, the mean absolute percentage error (MAPE) and Theil's Inequality Coefficient/Theil's U [Fuertes et al. 2009; Rapach and Wohar 2006] are used in order to evaluate the prediction performance of our method.

In Equations (10) and (11), A_g and F_g indicate the closing prices for the g-th actual value and forecast value of G observations, respectively, in the retrieved chart. This is calculated from the historical closing price HC_t at day-t by finding the most similar candlestick chart; HC_{t+f} is the closing price for the next$f$ days without the trading market suspends transactions:

$$\mathrm{MAPE} = \frac{1}{G}\sum\nolimits_{g=1}^{G} \frac{\left|F_g - A_g\right|}{A_g}, \tag{10}$$

$$\mathrm{U} = \sqrt{\frac{1}{G}\sum\nolimits_{g=1}^{G} (F_g - A_g)^2} \Bigg/ \sqrt{\frac{1}{G}\sum\nolimits_{g=1}^{G} (F_g)^2} + \sqrt{\frac{1}{G}\sum\nolimits_{g=1}^{G} (A_g)^2}, \tag{11}$$

where $1 \leq g \leq G$ and $\mathrm{Fg} = A_g \times \left[1 + \frac{(HC_{t+f} - HC_t)}{HC_t}\right]$.

Finally, the time-consuming complex (computational complexity) candlestick chart mining process is divided into three parts including: (1) calculating the feature vector (here named W for convenience) of the query image; (2) taking the similarity measure between W and every feature vector in the image database (H) one by one and then finding the N closest images according to these vectors; and (3) displaying the N most similar images. Therefore, the computational complexity of parts 1–2 is O(WH) and of part 3 is O(H), and the total time complexity is O(WH).

4. EXPERIMENTAL RESULTS

4.1. Dataset and Experimental Environment

In this study, the dataset used contains candlestick charts and a closing price file, which are based on trading information obtained from the Dow Jones Industrial Average Index (INDU) for the period between 1990 and 2011. Each experimental image (candlestick chart) includes T days of information and the resolution is 768×416.

In fact, the stock price is a reflection of investors' expectations of stock market earnings, thus the stocks' ups and downs are related to the business cycle. For example, if the investors have optimistic expectations about economic expansion, this means that investment will rise and employment opportunities and income will grow, resulting in a rise in stock prices. Conversely, if investors have a pessimistic view of the future economy, the stock prices will be expected to fall to reflect this expectation. In business cycle research [Groot and Franses 2012] the stock market cycles are called the Kitchin cycle (3–4 years), Juglar cycle (7–11 years), and Kuznets cycle (22–23 years). After World War II, the United States economy experienced a total of 10 cycles from 1945 to 2004, with the average length of every economic cycle from one trough to the next being 67 months (5.5 years). The longest economic cycle length was 128 months (10.5 years) from 1991 to 2001 [Colander 2012].

For these reasons, in this study, the dataset is divided into two subsets, 5-5 and 10-10, including: (1) five-year training (2000–2004) and five-year testing (2005–2009); and (2) ten-year training (1990–1999) and ten-year testing (2000–2009), respectively. The 5-5 dataset contains 1256 and 1259 training and testing charts and the 10-10 dataset contains 2528 and 2515 training and testing charts. Finally, we use the candlestick charts for INDU in 2010 (252 candlestick charts) and 2011 (253 candlestick charts) to further validate the prediction performance of our method.

In general, technical analysis is more suitable for short-term (one week to a month) and medium-term (one to three months) stock price forecasting. Since there are about 20 trading days in a month, in the pilot experiment, every candlestick chart was set to have T consecutive trading days (T = 20, 30, 40, 50, 60). The 5-5 dataset was then used with the 6×T feature to predict the next-day stock trend. The retrieval performance was 0.6311 (T = 20), 0.6823 (T = 30), 0.6612 (T = 40), 0.6043 (T = 50), and 0.5484 (T = 60). We found the optimal result for constructing a candlestick chart to be based on 30 consecutive trading days, thus T = 30 is used for the later experiments.

Table I. The Seven Extracted Feature Representations

Abbr.	Feature vector content	Dims.
30-A	AS_x	30
30-B	CBS_x	30
90	$AS_x + US_x + LS_x$	90
120-A	$AS_x + US_x + LS_x + CBS_x$	120
120-B	$US_x + LS_x + ULP_x + LLP_x$	120
150	$AS_x + US_x + LS_x + ULP_x + LLP_x$	150
180	$AS_x + US_x + LS_x + CBS_x + ULP_x + LLP_x$	180

In addition, using 30 consecutive trading days for a candlestick chart means that the 30 closing prices can be seen as different class labels (prediction outputs) with which to evaluate the retrieval performance. In other words, there are 30 different sets of (binary) class labels for predicting next-day stock movement corresponding to the 30 closing prices, in which each set of class labels represents either the movement of the stock price up or down. Therefore, for simplicity, we use the first (day-1), middle (day-15) and last (day-30) trading days as the basis of the class labels for stock prediction. Then, the retrieval performances obtained using these three different class labels are compared.

In order to find out which feature descriptor(s) are more representative and can provide better prediction performance, we extract seven different texture features for comparison as shown in Table I. In the first experiment in this study, the day-1, day-15 and day-30 (query) candlestick charts are used to predict the next 1, 7, 15, 30, and 90-day stock movements.

4.2. Experimental Results I

The aim of the first experiment is to find out which date is the best for the class label (day-1, day-15, or day-30) for predicting the stock price (from which to predict the next 1, 7, 15, 30, and 90-day's stock price). However, it should be noted that the retrieval results are obtained with our approach based on the r-th similar historical candlestick charts, and the retrieval results may differ when obtained using different numbers of similar historical candlestick charts. Thus, the voting method is used to predict the stock price from the r-th similar charts. For example, assume that the investor inputs a query candlestick chart and chooses g as the maximum voting value. If g is set to 5, the retrieval results of the top five similar charts are $r_1(-)$, $r_2(+)$, $r_3(+)$, $r_4(+)$, $r_5(-)$ for predicting the next day stock trend, where "+" and "–" indicate an increase or decrease in the stock price. The prediction results indicate that the stock price goes up after using the voting method. Since the 5-5 and 10-10 testing data sets are different (comprised of 1259 or 2515 candlestick charts), the maximum voting values of the 5-5 and 10-10 datasets are set to be 500 and 1000, respectively (the top 500 and 1000 similar charts) as part of the experimental parameters.

Figure 4 shows the results obtained using feature vector 180 in Table I and the 5-5 dataset by three different class labels. We can see that the performance obtained using class labels based on day-1 and day-15 is better than those for day-30, and the best voting intervals converge from 47 to 79 and 107 to 121 for the day-1 and day-15 class labels, respectively. Consequently, the median values of both class labels are chosen, which are 63 and 114, as the best voting values. Similarly, for the 10-10 dataset, the best voting intervals converge from 9 to 21 and 31 to 39 for the day-1 and day-15 class labels, respectively. Therefore, 15 is chosen as the voting value for day-1 and 29 for day-15.

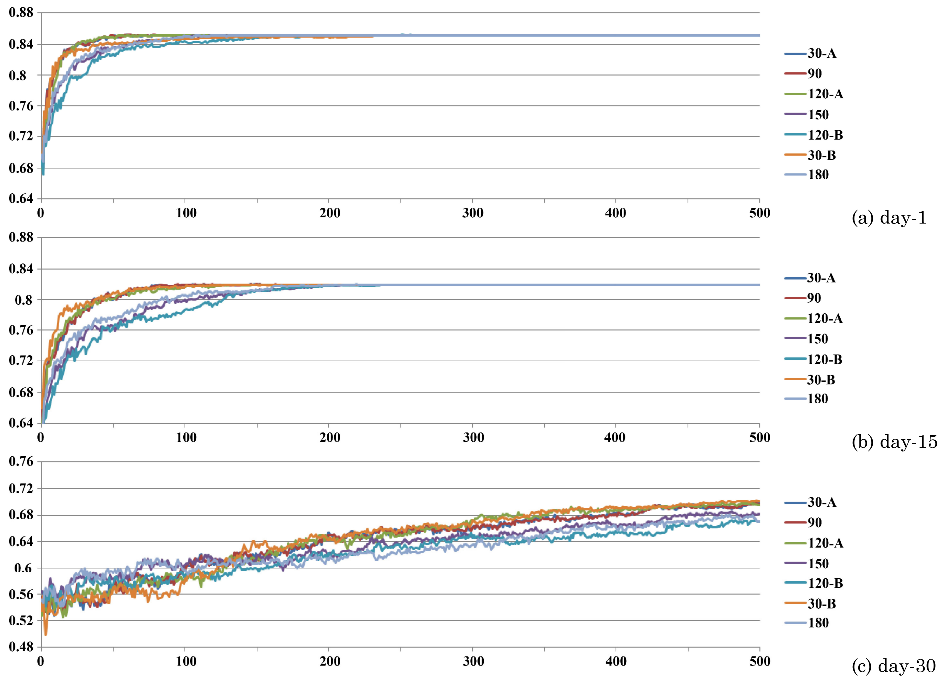


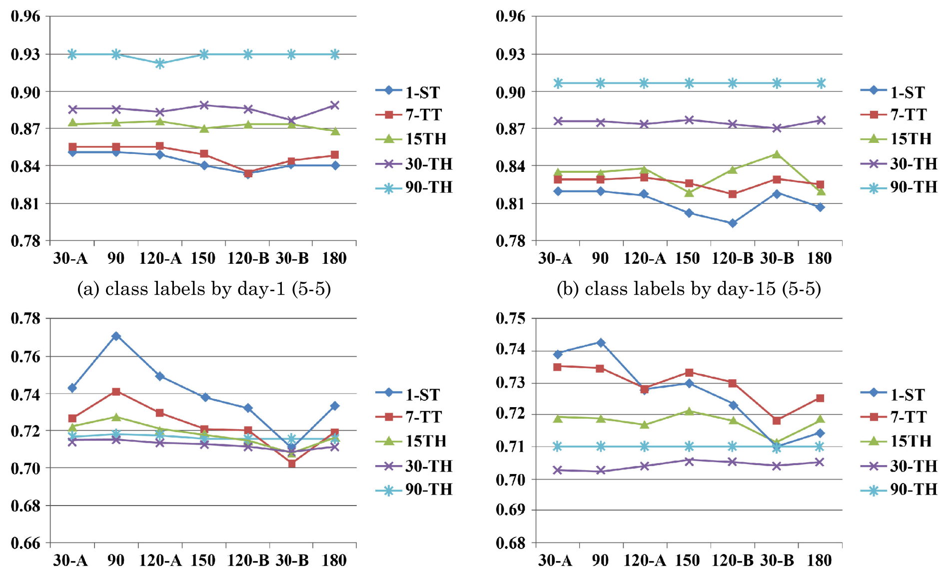
Fig. 4. The F-measure rates obtained using different feature vectors and different class labels.

Figure 5 shows the F-measure rate obtained using different class labels. Figures 5(a)–5(b) and Figures 5(c)–5(d) show the results for the 5-5 and 10-10 datasets. For the 5-5 dataset, the class label based on day-1 offers relatively better feature descriptor (or vector) performance compared with the class labels based on day-15. In particular, the feature descriptors for the 5-5 dataset can provide an F-measure of at least 0.794. However, these results are abnormal because the performance improves when the prediction period becomes longer.

On the other hand, the feature descriptors for the 10-10 dataset can provide an F-measure of 0.745, which is better than that of the 5-5 dataset (as shown in Figures 5(c)–5(d)). However, the F-measure gradually decreases as the prediction period becomes longer, which means that they are consistent with each other. Therefore, day-1 and day-15 for the 10-10 dataset are used in the following experiments.

4.3. Experimental Results II

In the second experiment, we examine the seven different feature vectors in terms of the MAPE and Theil's U rates. It can be seen by looking at Figures 6(a)–6(b) that the MAPE rates obtained using day-15 as the class label are better than with day-1, no matter what feature vectors are used. In addition, the Theil's U rates obtained using day-15 are also higher than for day-1, as shown in Figures 6(c)–6(d). Moreover, the MAPE rate is about 0.03 for predicting the next day's stock price, and the results are between 0.038–0.06 for the next 7-, 15-, and 30-day predictions (Figure 6(b)). On the other hand, the Theil's U rate is about 0.02 for predicting the next day's stock price, and the results are between 0.02–0.038 for the next 7-, 15-, and 30-day predictions (Figure 6(d)).



(a) class labels by day-1 (5-5)
(b) class labels by day-15 (5-5)
(c) class labels by day-1 (10-10)
(d) class labels by day-15 (10-10)

Fig. 5. The F-measure rates for different feature vectors and different class labels.

In particular, when the class label is based on day-15, the MAPE and Theil's U rates obtained using these feature vectors for the next 1-, 7-, 15-, and 30-day predictions are very similar. The Kruskal-Wallis (K-W) test [Corder and Foreman 2009] is used to verify whether there is a significant difference when using these features for the MAPE and Theil's rates; detailed information is shown in Table II. The degree of freedom of this test is 6 (7–1), and the test statistic is based on

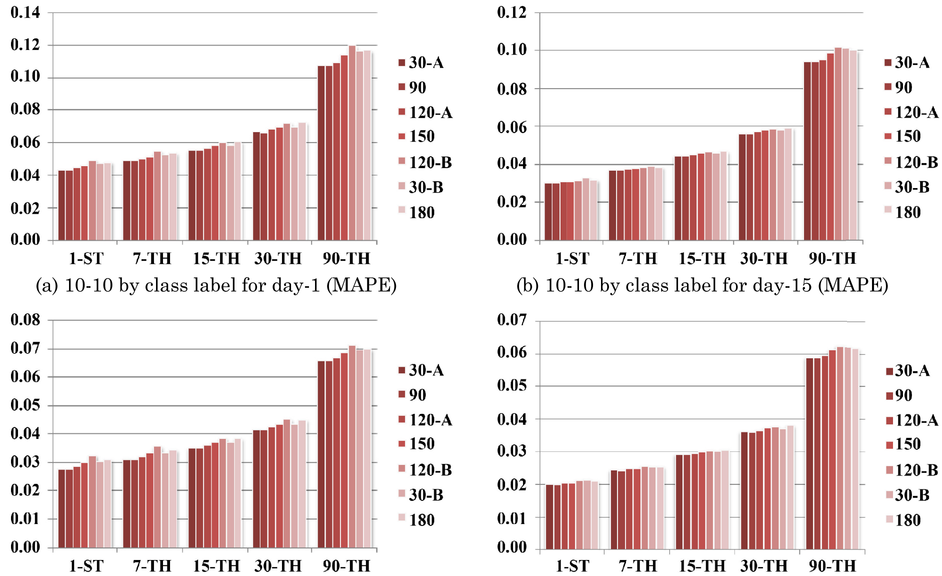
$$K = \frac{12}{n(n+1)} \sum_{i=1}^{C} \frac{R_i^2}{n_i} - 3(n+1), \tag{12}$$

where n is the total number of samples; c is the group number; and n_i and R_i are the number of the i-th group samples and the sum of the rank, respectively.

According to Table II, the test statistics are 1.227 and 1.188, respectively, with a confidence level of less than 95% (${\chi_{0.05}}^2(6) = 14.067$). Therefore, we can assert that there is no significant difference in the prediction performance obtained using these feature vectors. Hence, using the 10-10 dataset with the class labels based on day-15 (see Figure 6(b) and Figure 6(d)) can be regarded as the most representative experimental setting for comparing different models, which are investigated next. Furthermore, the performance obtained using feature vector 90 is better than with the other features, regardless of which evaluation methods are used. Therefore, we choose these two features for further discussion.

4.4. Comparison between Our Approach and Related Work

In the literature, some researchers have directly utilized trading data or technical indicators as the input variables to construct models for the prediction of values and trends in the Dow Jones Industrial Average Index (INDU) [Asadi et al. 2012; Hsieh et al. 2011; O'Connor and Madden 2006], and the selected input variables (technical



(a) 10-10 by class label for day-1 (MAPE)

(b) 10-10 by class label for day-15 (MAPE)

(c) 10-10 by class label for day-1 (Theil's U)

(d) 10-10 by class label for day-15 (Theil's U)

Fig. 6. Comparisons of the MAPE and Theil's U rates using different feature vectors.

Table II. MAPE and Theil's U Results Obtained Using Different Feature Vectors by Day-15

		30-A	90	120-A	150	120-B	30-B	180
1-st	MAPE	0.0304	0.0303*	0.0310	0.0309	0.0316	0.0328	0.0318
	Theil's U	0.0199	0.0199*	0.0203	0.0204	0.02 11	0.0213	0.0209
7-th	MAPE	0.0373	0.0372*	0.0378	0.0379	0.0387	0.0389	0.0388
	Theil's U	0.0243	0.0242*	0.0247	0.0248	0.0255	0.0254	0.0253
15-th	MAPE	0.0448	0.0447*	0.0454	0.0460	0.0465	0.0462	0.0471
	Theil's U	0.0292	0.0291*	0.0295	0.0299	0.0303	0.0301	0.0305
30-th	MAPE	0.0566	0.0565*	0.0573	0.0582	0.0589	0.0584	0.0596
	Theil's U	0.0360	0.0360*	0.0365	0.0372	0.0376	0.0370	0.0379
90-th	MAPE	0.0946	0.0945*	0.0956	0.0990	0.1021	0.1017	0.1004
	Theil's U	0.0588	0.0588*	0.0595	0.0610	0.0623	0.0620	0.0616

indicators) for stock prediction. For simplicity, we call TA_1 (31 variables[1] in O'Connor and Madden [2006]), TA_2 (9 variables[2] in Hsieh et al. [2011]), and TA_3 (7 variables[3] in Asadi et al. [2012]), and they are compared with our approach over the 10-10 dataset.

The stock price trend is an important signal for stock investors, and can be treated as a binary classification problem, *i.e.*, stock price ups and downs. In general, there

[1] TA_1: current day's INDU opening value, previous 5 days' INDU opening values, previous 5 days' INDU gradients, previous 5 days' WTI Cushing crude oil price, previous 5 days of the USD/YEN exchange rate, previous 5 days of the USD/GBP exchange rate, and previous 5 days of the USD/CAN exchange rate.

[2] TA_2: 6-day moving average (MA_6), demand index (DI), 12-day exponential moving average (EMA_12), 26-day exponential moving average (EMA_26), moving average convergence/divergence (MACD), the close price one day ago (C), the open price one day ago (O), the highest price one day ago (H), the lowest price one day ago (L).

[3] TA_3: 6-day moving average (MA_6), 6-day bias (BIAS_6), 6-day relative strength index (RSI_6), 9-day stochastic oscillator (KD_9), moving average convergence/divergence (MACD), 13-day psychological line (PSY), and trading volume (VOL).

Table III. Comparison between Different Methods (Accuracy/Precision/Recall/F-Measure)

	k-NN	SVM	BPN
TA_1	0.603/0.695/0.629/0.661	0.546/0.967/0.547/0.699	0.582/0.908/0.581/0.709
TA_2	0.552/0.975/0.553/0.706	0.570/0.977/0.568/0.718	0.616/0.865/0.615/0.719
TA_3	0.559/0.929/0.564/0.702	0.606/0.832/0.613/0.706	0.639/0.871/0.626/0.729
Our method	0.634/0.951/0.610/0.743		

Table IV. Comparison between Other Methods and Our Method (MAPE Rates/Theil's U Rates)

	GARCH (1, 1)	SVR	BPN
TA_1	0.390/0.244	0.159/0.108	0.036/0.028
TA_2	0.355/0.227	0.119/0.080	0.032/0.022
TA_3	0.362/0.291	0.172/0.097	0.048/0.031
Our method	0.0303/0.0199		

are several related techniques that can be used for dealing with this kind of problem, including the decision tree (DT), k-nearest neighbor (k-NN), and SVM models. The ANN is one of the most well-known techniques, also being widely used for many other classification problems. Here, the k-NN model and SVM model are constructed with the radial based function (RBF) kernel using a cost value of $c = 2^8$ and a gamma value of $\gamma = 2^{-7}$. The backpropagation network (BPN) model uses a learning rate of $\eta = 0.05$ and the momentum coefficient $\alpha = 0.5$ in the hidden layer for a further comparison. Table III shows the prediction performance of the next-day stock price trend obtained with the k-NN, SVM, and BPN models using different input variables. Feature vector 90 using day-15 as the class label is used in our proposed method. Compared with other prediction models, our model demonstrates better performance, and our similarity measure approach also has the advantage that it does not require model training but can still provide relatively good performance.

In addition, it should be noted that prediction of the closing price can also be regarded as a time series prediction problem. There are many related models for solving this type of problem (a regression problem), including statistical methods (e.g., ARIMA and GARCH) and machine learning methods (e.g., support vector regression (SVR) and BPN). The different technical indicators are used as the input variables to train the SVR and BPN models with the same parameters. Table IV shows the performance results for predicting the closing price from testing charts, finding the MAPE and Theil's U rates by using the GARCH (1, 1), SVR, and BPN methods. Clearly, the MAPE and Theil's U rates obtained with our method (using the 90 feature vector as the class label and day-15) are lower than those obtained with the other models that use different technical indicators: approximately 3% and 2%, respectively. These results indicate that our candlestick chart analysis method is much more suitable for stock prediction than the traditional prediction models.

4.5. Validation

Finally, the 2010 (2011) dataset, which contains 252 (253) candlestick charts is used to validate the performance of the proposed method in terms of the accuracy, precision, recall, and F-measure for binary classification, and the MAPE and Theil's U rates for predicting the closing price.

It can be seen in the left part (2010 dataset) of Table V that the accuracy and precision rates are all over 0.64, and the F-measure rates are over 0.78 for the next 1-, 7- and 15-day predictions. Moreover, the recall rate exceeds 0.98, which means that

Table V. Comparison between Methods

	2010			2011		
Prediction rate	1-st	7-th	15-th	1-st	7-th	15-th
Accuracy	72.61%	65.02%	64.90%	68.77%	67.98%	69.96%
Precision	73.81%	65.48%	64.68%	71.02%	69.74%	68.75%
Recall	98.24%	98.75%	98.15%	81.70%	86.08%	95.65%
F-measure	83.50%	78.41%	78.13%	75.99%	77.05%	80.00%
Prediction performance	1-st	7-th	15-th	1-st	7-th	15-th
MAPE	2.625%	3.157%	3.692%	2.945%	3.361%	3.694%
Theil's U	1.662%	1.958%	2.250%	1.972%	2.235%	2.451%

the proposed method provide a very effective indication of investment direction when the retrieval results indicate a bull market. As a result, we can say that candlestick chart mining is very suitable for active investors. Specifically, the texture features provide about 72.61% accuracy, 73.81% precision, 98.24% recall, and 83.50% F-measure for predicting the next day stock movement based on a given query chart, while the prediction performances using feature vector 90 are 2.625% for the MAPE and 1.662% for the Theil's U rates for next-day prediction. On the right side (2011 dataset) of Table V, it can be seen that the texture features can provide about 68.77% accuracy, 71.02% precision, 81.70% recall, and 75.99% F-measure for predicting the next day stock movement based on a given query chart, while the prediction performance using feature vector 90 is 2.945% for the MAPE and 1.972% for the Theil's U rates for next-day prediction.

Note that we also examine the applicability of our proposed method to the real market in real time, by making predictions for the period from March 20, 2013 to May 1, 2013 using the candlestick chart of March 19, 2013 as the query, where its base value is 14,455.82. The prediction results, without the suspension of trading because of holidays are as follows: 2013/03/20: 14,512.99 (+), 2013/03/28: 14,278.66 (–), 2013/04/10: 14,337.81 (–), and 2013/05/01: 14,612.57 (+).

5. CONCLUSION AND DISCUSSION

In this paper, we present a new candlestick chart analysis method for making stock predictions based on the idea of content-based image retrieval (CBIR). This method is composed of feature extraction, image retrieval, and knowledge discovery steps. The feature extraction step is based on extracting the wavelet texture features from candlestick charts, where seven different texture features are extracted for comparison. In the image retrieval step, the Euclidean distance is used to measure the similarity among candlestick charts for retrieval. In the final step, the retrieval results are evaluated to predict stock movement.

Experiments are carried out based on two different datasets as well as one validation dataset collected from the Dow Jones Industrial Average Index (INDU). The results show that using the 10-10 dataset (10-years for training and testing) with the class labels collected from day-15 of the candlestick charts allows for better performance of the different wavelet based texture feature representations than with the other dataset settings, such as the 5-5 dataset (5-years for training and testing) and the class labels for the day-1 and day-30 candlestick charts, respectively. In addition, we found that the 30-A, 90, and 120-A feature vectors are more suitable for predicting stock movements, with the best short- and medium-term stock movement performance being obtained for the 90 feature vector. More significantly, comparison between the

30-A, 30-B, 90, and 120-A feature vectors shows that retrieval performance for short- and medium-term stock investment is the worst for the 30-B feature vector, and the difference between the 90 and 120-A feature vectors is the middle point of the candle bodies' pixel values (30-B). Thus we conclude that the 30-B feature vector is not suitable for stock prediction. It produces a 3.031% MAPE and 1.988% Theil's U rates for the 10-10 dataset and 2.625% MAPE in 2010 (2.945% MAPE in 2011) and 1.622% Theil's U rates in 2010 (1.972% Theil's U rates in 2011) for the validation datasets when predicting the next-day stock price. In addition, prediction performance is better for the group with more rather than fewer participants when choosing 15 and 29 for day-1 and day-15, respectively.

We can conclude that the visual content of candlestick charts is indeed useful for stock prediction. In particular, candlestick chart analysis by texture feature extraction and similarity matching can provide reasonably good prediction performance in terms of precision/recall, F-measure, MAPE, and Theil's U. In other words, the content-based image retrieval technique is applicable to stock prediction.

Several issues remain to be considered in the future. The first is to extract other (novel) visual features of candlestick charts, such as the shape and/or edges, to improve prediction performance. Second, different types of features can be combined, such as the traditional fundamental and technical indices and visual features for further examination. Third, the next issue would be to assess the prediction performance of this candlestick chart analysis method in different stock markets. Last but not least, the INDU's components have changed many times since its inception, and this is also an important issue for future research.

REFERENCES

Andreopoulos, Y. 2009. Comments on "phase-shifting for nonseparable 2-D Haar wavelets. *IEEE Trans. Image Process. 18*, 8, 1897–1898.

Araújo, R. D. A. 2010. Hybrid intelligent methodology to design translation invariant morphological operators for Brazilian stock market prediction. *Neural Netw. 23*, 10, 1238–1251.

Araújo, R. D. A. and Ferreira, T. A. E. 2013. A morphological-rank-linear evolutionary method for stock market prediction. *Inform. Sci., 237*, 3–17.

Ariss, R. T., Rezvanian, R., and Mehdian, S. M. 2011. Calendar anomalies in the Gulf Cooperation Council stock markets. *Emerg. Markets Rev. 12*, 3, 293–307.

Asadi, S., Hadavandi, E., Mehmanpazir, F., and Nakhostin, M. M. 2012. Hybridization of evolutionary Levenberg–Marquardt neural networks and data pre-processing for stock market prediction. *Knowl.-Based Syst. 35*, 245–258.

Bollerslev, T. 1986. Generalized autoregressive conditional heteroscedasticity. *J. Econometrics 31*, 3, 307–327.

Briza, A. C. and Naval Jr., P. C. 2011. Stock trading system based on the multi-objective particle swarm optimization of technical indicators on end-of-day market data. *Appl. Soft Comput. 11*, 1, 1191–1201.

Brock, W., Lakonishok, J., and LeBaron, B. 1992. Simple technical trading rules and the stochastic properties of stock returns. *J. Finance 47*, 5, 1731–1764.

Caginalp, G. and Laurent, H. 1998. The predictive power of price patterns. *Appl. Math. Finance 5*, 3–4, 181–205.

Chan, S. W. K. and Franklin, J. 2011. A text-based decision support system for financial sequence prediction. *Decision Supp. Syst. 52*, 1, 189–198.

Chang, P. H. K. and Osler, C. L. 1999. Methodical madness: Technical analysis and the irrationality of exchange-rate forecasts. *Econ. J. 109*, 458, 636–661.

Chen, D. R., Huang, Y. L., and Lin, S. H. 2011. Computer-aided diagnosis with textural features for breast lesions in sonograms. *Comput. Med. Imaging Graphics 35*, 3, 220–226.

Chen, P. C. and Pavlidis, T. 1983. Segmentation by texture using correlation. *IEEE Trans. Pattern Anal. Mach. Intell. 5*, 1, 64–69.

Cheng, C. H., Chen, T. L., and Wei, L. Y. 2010. A hybrid model based on rough sets theory and genetic algorithms for stock price forecasting. *Inform. Sci. 180*, 9, 1610–1629.

Chianga, J. Y. and Cheng, S. R. 2009. Multiple-instance content-based image retrieval employing isometric embedded similarity measure. *Pattern Recog. 42*, 1, 158–166.

Colander, D. C. 2012. *Macroeconomics* 9th Ed. McGraw-Hill, New York, NY.

Corder, G. W. and Foreman, D. I. 2009. *Nonparametric Statistics for Non-Statisticians: A Step-By-Step Approach* 1st Ed. Wiley, New York, NY.

de Groot, B. and Franses, P. H. 2012. Common socio-economic cycle periods. *Technol. Forecasting Soc. Change 79*, 1, 59–68.

Doyle, J. R. and Chen, C. H. 2009. The wandering weekday effect in major stock markets. *J. Banking Finance 33*, 8, 1388–1399.

Edwards, R. D., Magee, J., and Bassetti, W. H. C. 2012. *Technical Analysis of Stock Trends* 10th Ed. CRC Press, Boca Raton, FL.

Engle, R. F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica 50*, 4, 987–1007.

Fama, E. F. 1970. Efficient capital markets: A review of theory and empirical work. *J. Finance 25*, 2, 383–417.

Fiess, N. M. and MacDonald, R. 2002. Towards the fundamentals of technical analysis: Analysing the information content of high, low and close prices. *Econ. Modell. 19*, 3, 353–374.

Fock, J. H., Klein, C., and Zwergel, B. 2005. Performance of candlestick analysis on intraday futures data. *J. Derivatives 13*, 1, 28–40.

Friesen, G. C., Weller, P. A., and Dunham, L. M. 2009. Price trends and patterns in technical analysis: A theoretical and empirical examination. *J. Banking Finance 33*, 6, 1089–1100.

Fuertes, A. M., Izzeldin, M., and Kalotychou, E. 2009. On forecasting daily stock volatility: The role of intraday information and market conditions. *Int. J. Forecasting 25*, 2, 259–281.

Graña, M. and Veganzones, M. A. 2012. An endmember-based distance for content based hyperspectral image retrieval. *Pattern Recog. 45*, 9, 3472–3489.

Hagenau, M., Liebmann, M., and Neumann, D. 2013. Automated news reading: Stock price prediction based on financial news using context-capturing features. *Decision Supp. Syst. 55*, 3, 685–697.

Hansen, P. R. and Lunde, A. 2005. A forecast comparison of volatility models: Does anything beat a GARCH(1,1)? *J. Appl. Econometrics 20*, 7, 873–889.

He, D. C. and Wang, L. 1990. Texture unit, texture spectrum, and texture analysis. *IEEE Trans. Geosci. Remote Sens. 28*, 4, 509–512.

Hsia, C. H., Chiang, J. S., and Guo, J. M. 2013. Memory-efficient hardware architecture of 2-D dual-mode lifting-based discrete wavelet transforms. *IEEE Trans. Circuits Syst. Video Technol. 23*, 4, 671–683.

Hsia, C. H., Guo, J. M., and Chiang, J. S. 2009. Improved low-complexity algorithm for 2-D integer lifting-based discrete wavelet transform using symmetric mask-based scheme. *IEEE Trans. Circuits Syst. Video Technol. 19*, 8, 1202–1208.

Hsieh, T. J., Hsiao, H. F., and Yeh, W. C. 2011. Forecasting stock markets using wavelet transforms and recurrent neural networks: An integrated system based on a artificial bee colony algorithm. *Appl. Soft Comput. 11*, 2, 2510–2525.

Hsu, C. M. 2011. A hybrid procedure for stock price prediction by integrating self-organizing map and genetic programming. *Expert Syst. Appl. 38*, 11, 14026–14036.

Hsu, P. H. and Kuan, C. M. 2005. Reexamining the profitability of technical analysis with data snooping checks. *J. Financial Econ. 3*, 4, 606–628.

Hung, J. C. 2009. A fuzzy asymmetric GARCH model applied to stock markets. *Inform. Sci. 179*, 22, 3930–3943.

Kang, M. 2010. Probability of information-based trading and the January effect. *J. Banking Finance 34*, 12, 2985–2994.

Krishnamachari, S. and Chellappa, R. 1997. Multiresolution Gauss-Markov random field models for texture segmentation. *IEEE Trans. Image Proc. 6*, 2, 251–267.

Kwon, Y. K. and Moon, B. R. 2007. A hybrid neurogenetic approach for stock forecasting. *IEEE Trans. Neural Netw. 18*, 3, 851–864.

Lee, C. H. L., Liu, A., and Chen, W. S. 2006. Pattern discovery of fuzzy time series for financial prediction employing an isometric embedded similarity measure. *Pattern Recog*. *42*, 1, 158–166.

Lee, C. H. L., Liu, A., and Chen, W. S. 2006. Pattern discovery of fuzzy time series for financial prediction. *IEEE Trans. Knowl. Data Eng. 18*, 5, 613–625.

Lee, D. J., Antani, S., Chang, Y., Gledhill, K., Long, L. R., and Christensen, P. 2009. CBIR of spine X-ray images on inter-vertebral disc space and shape profiles using feature ranking and voting consensus. *Data Knowl. Eng. 68*, 12, 1359–1369.

Lee, K. H. and Jo, G. S. 1999. Expert system for predicting stock market timing using a candlestick chart. *Expert Syst. Appl. 16*, 4, 357–364.

Leigh, W., Modani, N., Purvis, R., and Roberts, T. 2002. Stock market trading rule discovery using technical charting heuristics. *Expert Syst. Appl. 23*, 2, 155–159.

Levy, T. and Yagil, J. 2012. The week-of-the-year effect: Evidence from around the globe. *J. Banking Finance 36*, 7, 1963–1974.

Lilly, J. M. and Olhede, S. C. 2010. On the analytic wavelet transform. *IEEE Trans. Inf. Theory 56*, 8, 4135–4156.

Lin, M. C., Lee, A. J. T., Kao, R. T., and Chen, K. T. 2011. Stock price movement prediction using representative prototypes of financial reports. *ACM Trans. Manage. Inf. Syst. 2*, 3, 11–18.

Liu, C. F., Yeh, C. Y., and Lee, S. J. 2012. Application of type-2 neuro-fuzzy modeling in stock price prediction. *Appl. Soft Comput. 12*, 4, 1348–1358.

Liu, F. and Wang, J. 2012. Fluctuation prediction of stock market index by Legendre neural network with random time strength function. *Neurocomput. 83*, 12–21.

Luo, L. and Chen, X. 2013. Integrating piecewise linear representation and weighted support vector machine for stock trading signal prediction. *Appl. Soft Comput. 13*, 2, 806–816.

Marshall, B. R., Young, M. R., and Cahan, R. 2008. Are candlestick technical trading strategies profitable in the Japanese equity market? *Rev. Quant. Finance Account. 31*, 2, 191–207.

Marshall, B. R., Young, M. R., and Rose, L. C. 2006. Candlestick technical trading strategies: Can they create value for investors. *J. Banking Finance 30*, 8, 2303–2323.

Moller, N. and Zilca, S. 2008. The evolution of the January effect. *J. Banking Finance 32*, 3, 447–457.

Morris, G. L. 2006. *Candlestick Charting Explained: Timeless Techniques for Trading Stocks and Futures* 3rd Ed. McGraw-Hill, New York, NY.

Murphy, J. J. 1999. *Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications* 1st Ed. New York Institute of Finance, New York, NY.

O'Connor, N. and Madden, M. G. 2006. A neural network approach to predicting stock exchange movements using external factors. *Knowl.-Based Syst. 19*, 5, 371–378.

Oh, S. K., Pedrycz, W., and Park, H. S. 2006. Genetically optimized fuzzy polynomial neural networks. *IEEE Trans. Fuzzy Syst. 14*, 1, 125–144.

Park, K. and Shin, H. 2013. Stock price prediction based on a complex interrelation network of economic factors. *Eng. Appl. Artif. Intell. 26*, 5–6, 1550–1561.

Qi, M. and Zhang, G. P. 2008. Trend time-series modeling and forecasting with neural networks. *IEEE Trans. Neural Netw. 19*, 5, 808–816.

Quellec, G., Lamard, M., Cazuguel, G., Cochener, B., and Roux, C. 2010. Adaptive nonseparable wavelet transform via lifting and its application to content-based image retrieval. *IEEE Trans. Image Proc. 19*, 1, 25–35.

Rapach, D. E. and Wohar, M. E. 2006. In-sample vs. out-of-sample tests of stock return predictability in the context of data mining. *J. Empirical Finance 13*, 2, 231–247.

Rui, Y., Huang, T. S., and Chang, S. F. 1999. Image retrieval: Current techniques, promising directions and open issues. *J. Visual Commun. Image Rep. 10*, 1, 39–62.

Sapankevych, N. I. and Sankar, R. 2009. Time series prediction using support vector machines: A survey. *IEEE Comput. Intell. Mag. 4*, 2, 24–38.

Schumaker, R. P. and Chen, H. 2009. Textual analysis of stock market prediction using breaking financial news: The AZFinText system. *ACM Trans. Inf. Syst. 27*, 2, 1–19.

Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., and Jain, R. 2000. Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Anal. Mach. Intell. 22*, 12, 1349–1380.

Tsai, C. F. and Hsiao, Y. C. 2010. Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches. *Decis. Supp. Syst. 50*, 1, 258–269.

Wang, Y. F., Cheng, S., and Hsu, M. H. 2010. Incorporating the Markov chain concept into fuzzy stochastic prediction of stock indexes. *Appl. Soft Comput. 10*, 2, 613–617.

Wink, A. M. and Roerdink, J. B. T. M. 2010. Polyphase decompositions and shift-invariant discrete wavelet transforms in the frequency domain. *Signal Proc. 90*, 6, 1779–1787.

Yue, J., Li, Z., Liu, L., and Fu, Z. 2011. Content-based image retrieval using color and texture fused features. *Math. Comput. Modell. 54*, 3–4, 1121–1127.

Zarandi, M. H. F., Zarinbal, M., Ghanbari, N., and Turksen, I. B. 2013. A new fuzzy functions model tuned by hybridizing imperialist competitive algorithm and simulated annealing application: Stock price prediction. *Inf. Sci. 222*, 213–228.

Zarb, F. G. and Kerekes, G. T. 1970. *The Stock Market Handbook: Reference Manual for the Securities Industry* 1st Ed. Dow Jones-Irwin, Homewood, IL.

Zheng, P. and Huang, J. 2013. Discrete wavelet transform and data expansion reduction in homomorphic encrypted domains. *IEEE Trans. Image Proc. 22*, 6, 2455–2468.

Zhu, Y. and Zhou, G. 2009. Technical analysis: An asset allocation perspective on the use of moving averages. *J. Financial Econ. 92*, 3, 519–544.

Received March 2013; revised October 2013; accepted February 2014