

Monitoring of Humans Traffic using Hierarchical Temporal Memory Algorithms

S. Sinkevicius, R. Simutis, V. Raudonis

Institute of Automation and Control Systems (AVATI), Kaunas University of Technology,

Faculty of Electrical and Control Engineering, Kaunas University of Technology,

Studentų str. 48-327, Kaunas LT-51367, Lithuania, +370 37 300296, e-mails: saulius@sinkevicius.net,

rimvydas.simutis@ktu.lt

crossref <http://dx.doi.org/10.5755/j01.eee.115.9.757>

Introduction

This paper explores possibilities of using Hierarchical Temporal Memory (HTM) [1] mechanism to calculate the humans' traffic balance in a public places, like rooms, trolleybuses, buses, shops, etc. This information can be used for tracking human-traffic bottlenecks, reducing loads, distributing traffic and adjusting business processes according to human's activity.

In recent years there have been many interesting researches and developments concerning to this problem. For human detection and classification tasks various modification of support vector machines (SVM) algorithms [2], human body kinematics [3], shape matching as well as features extraction algorithms and artificial neural networks are used [4]. Furthermore edge-detection algorithms [5], optical flow methods [6], weight map algorithms [7], and motion analysis [8] were applied for human detection tasks. Universality in different context is acquired with background extraction, calculating frame difference or using some sort of filters like Kalman [9] or Gaussian.

None of those methods can handle all issues of problems connected with calculation of humans' traffic in public places. Most difficult situations arise, when lighting changes rapidly, the place is crowded or humans visually merges and so on. In those situations described methods are very inaccurate, because they don't have any memory of previews frames and can't adequately react to changing environment. To solve these problems we choose an approach, based on HTM algorithms [10].

Recently, innovative IT company Numenta released a very promising framework [11, 12] for a vision cognition tasks, but some essential algorithms are not implemented yet. In this paper we propose a special extension and alternative use of Numenta methods.

In our application the cognitive mechanism is implemented by recognizing when a human appears in a particular area of the frame and identifying the direction of

the movement. With those information elements the proposed algorithm calculates balance of humans' traffic.

Two HTM networks were trained for different purposes. One instance of HTM network recognizes humans in a particular area of frame and another one - identifies direction of movement.

Theoretical background

HTM is a memory system that implements the structural and algorithmic properties of the neocortex [13, 14]. Instead of programming individual solutions for each problem, this method uses common algorithms to solve many cognitive tasks.

One of basic HTM functions is to discover causes. Anything that interacts with our senses like physical objects or even abstract things like ideas can be expressed as patterns. And those patterns always have causes. So, every input pattern is related to concrete causes (Figure 1, a). A belief is the moment-to-moment distribution of possible causes.

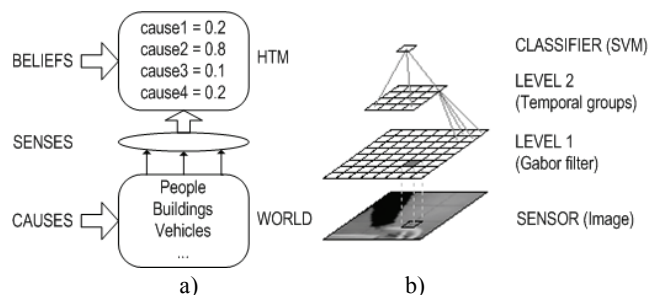


Fig. 1. HTM concept (a) [1] and three levels of the HTM network (b)

By applying the HTM the spatial-temporal aspects are essential. A spatial property allows creating a finite alphabet of input patterns and a temporal property allows relate input variations in time to same causes. A HTM

network is trained using continuously changing data over time while causes remain relatively stable.

Network is organized as tree shaped hierarchy of nodes and each node executes the same algorithms (see Fig. 1, b). Learning data are exposed to a sensor; which is at the bottom of network. Then, the information flows up through the network hierarchy to the top of the network and at each network-level more abstract beliefs are formed. In a learning phase, value of category goes directly to the top node (a supervised learning). By the time, when data comes to the top node, it's already known to what category the input belongs. In the end of training procedure the hierarchical model of the world is formed. When all levels of the network are trained, we can apply this network to patterns recognition tasks (inference).

There are two types of inference: Flash Inference (FI) (we used “maxProp” mode which computes a more peaked score for the group based on the current input only) and Time Based Inference (TBI) [15]. TBI can help the network to produce better inference results where each successive input is temporally related to the previous one, because the output is computed based on the current and previous inputs. The system takes into account the likelihood of each coincidence to follow another (based on the order of coincidences seen during training) and uses this information to compute the output probabilities after each input. For example, if you have an image application and you are feeding in successive frames of a movie, TBI mode will produce better inference results than FI. The FI is best suited for applications where each successive inference input is completely independent from the previous one.

Experimental setup

The proposed human traffic counting system contains four phases (see Fig. 2). In the first phase frames are prepared for processing. In second humans are differentiated from other objects using first HTM network. In third phase, activity of several stripes of the segment is calculated. This activity is used to identify movement direction. And in fourth phase, using movement direction and human presenting probabilities, the balance can be calculated.

Observation setup and preprocessing in phase 1

Humans tracking in a public places starts from picking a good position for a web camera. The easiest way to record all moving objects through some passageway is to do it from above the passage-way. Then everything what moves below is clearly visible. Trying to record from other directions than top is worse, because some objects can hide each other. In our application data are frames, acquired from video camera placed above the doors as shown in Fig. 3. In this application video camera with 640x480 pixels resolution is used.

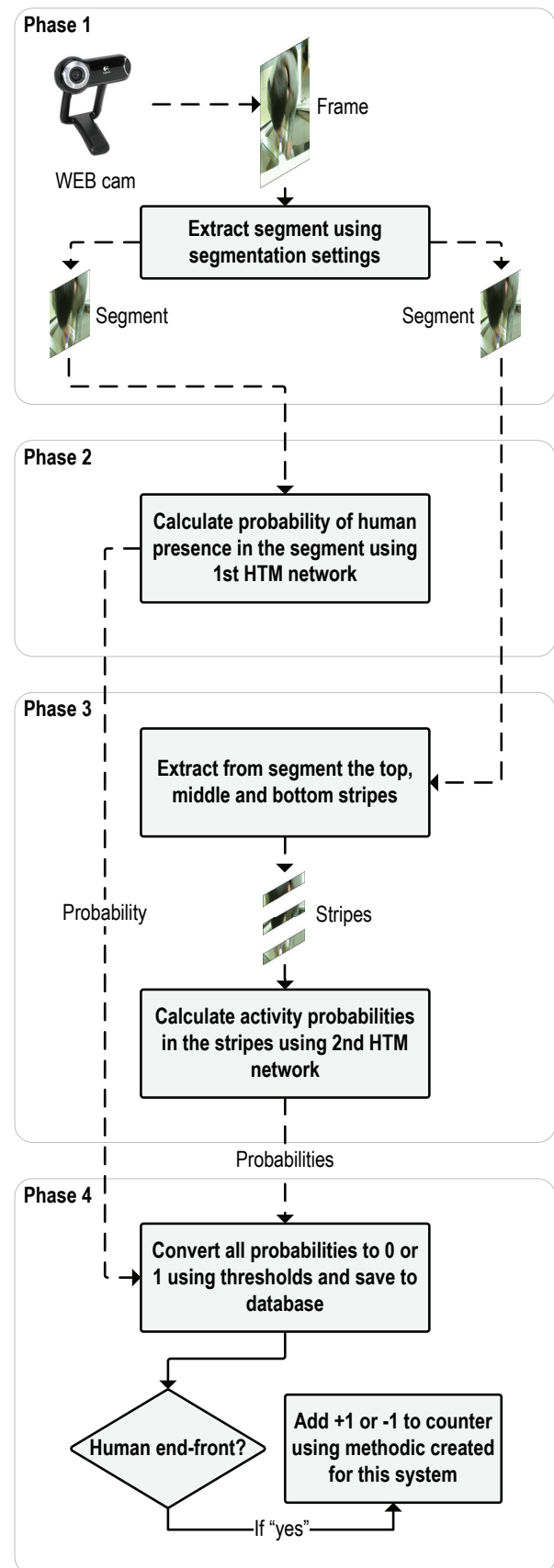


Fig. 2. Schema of the human traffic counting system

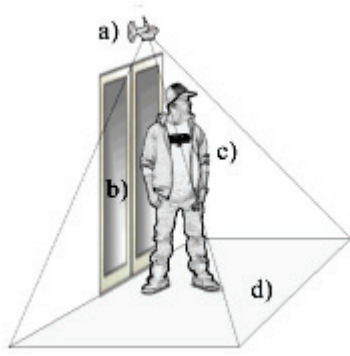


Fig. 3. The observation scheme. Web camera (a) is installed above the doors (b). Visible area is a square (d). Humans (c) enters or leaves a building through the door (b)

Doors differs from each other mostly only by wide. It's hard to say what distribution of wideness is, but definitely doors can be sorted by how many people can pass it in the same moment without collision. In most cases this number can vary from 1 to 3. The system was tested with narrow doors like for 1-2 people, but algorithm should still work with wider passage-ways but additional segments should be used.

Before starting the system we define area-segment in the whole frame which will be tracked (see Fig. 4). In our system this area is in the middle of doors with proportions of 2x1 and 480x230 size in pixels.

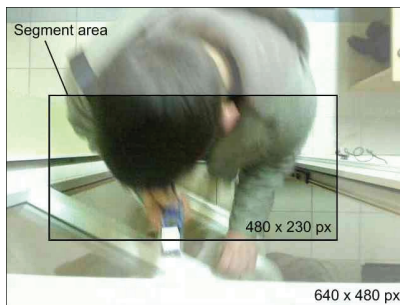


Fig. 4. An example of segmentation is shown. Size and location of a segment should be tuned by context specifics. We placed it in the middle of doors

Designing an HTM network

Numenta provided some tools to experiment with HTM algorithms. The core is NuPIC [12] framework based on Python programming language. It allows creating a HTM network in any structure, to train, to debug and to run inference on data sets after training is done. Visio Framework is an extension of NuPIC to design networks for vision problems only in a parameterized way without coding python. Vision Toolkit [16] has GUI, 2 optimized HTM networks for universal vision problems and let's to run experiments in only few clicks. All these tools were used for experimenting and some comparisons of accuracy and performance are shown in results section.

The biggest challenge is to design a HTM network which is fast enough, accurate and universal. Firstly we tried to create networks with Vision toolkit. This toolkit analyses data and choose one of two already prepared networks. Those networks are optimized, but they are too

big for our problem, slower than necessary and can't be used with TBI ability. It's why we decided to create own network.

Networks designed by Numenta are too large, so we designed own network of only 3 levels. Most parameters were tuned empirically, others taking into account desirable outcome. Here are some adjustments explained.

Sensor dimensions are 200x60 pixels in size. We found that this size is small enough to get reasonable good performance and is big enough not to lose important information.

We use grayscale pictures, because HTM can't deal with colors yet and it converts them to grayscale anyway.

First experiments was done using spatial pooler in the first level, but we found what it creates different outputs even if training set doesn't changes and this can lead to unpredictable results, because each time different spatial alphabet is formed. So, we have used NuPIC instrument "GaborNode" witch works like Gabor filter and always has same results with same input.

After network is designed, it should be trained in the proper way. Our data are temporal, so we should train the network in the sequence of movie.

HTM network for humans' detection (Phase 2)

We used 1358 real situations when humans enter or leave the room for training and testing an HTM network. Most situations are really complicated, because persons passes fast with tiny time spans in between or merges visually. The video is few hours long and the light is changing rapidly.

About 7000 pictures from different parts of video were used for training and are divided in two categories (Fig. 5).

Categorization of training data was empirical. We reviewed all training frames and classified them into two categories. The frames where big parts of humans are visible were assigned to the first category and other frames to the second category.

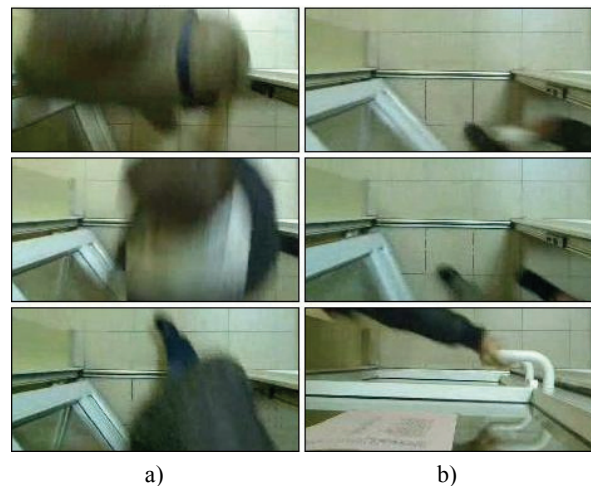


Fig. 5. A few examples from training data sets: a – frames with humans; b – others frames

Human's detection was done simply by analyzing one by one all segments in the testing data set. The HTM

network each time provides a probability between zero and one. Probability one is obtained for the cases when a human appears in the segment.

HTM network for movement detection (Phase 3)

When a human in the segment is detected the second task is to determine the movement direction of this human. If only one human pass the doors, it's easy to calculate the movement direction using various methods [6,8]. But this is a complicated task when several humans create one long moving object without enough time spans in between. In our analysis we have assumed that only one human could be in one segment in the particular moment of the time and all humans move from segment's top to bottom or from bottom to top. We put some small "zones" in the segment and then analyzed the activation sequences of these zones. The HTM network for movement detection must detect objects in the "zone". We placed three such activity monitoring "zones" in the top, middle and bottom of the segment. They are implemented by cutting 3 stripes from a segment (see Fig. 6) and calculating activity in each strip by HTM network for movement detection.

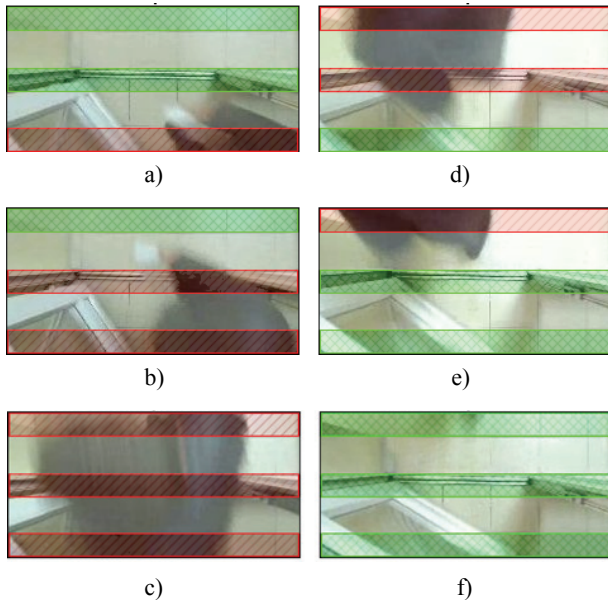


Fig. 6. An example of stripes activity when a human enters room (diagonal cross– idle, backward diagonal– active)

To detect humans' movement the sequence of stripes' activity peaks should be analyzed. When human enters the room, the bottom stripe is most active at first, then the middle stripe is activated and lastly the top stripe is active. If the sequence is otherwise: top, middle, bottom, we can state, that one human exited the room.

In more complex situations without time spans between humans, we could only look at first and last human assuming that all of them go in the same direction.

Parameters of the HTM network for movement detection are the same as for the first HTM (HTM for human detection) except the size of sensor. We use stripes

in 230x20 pixels in size. Learning strategy was also the same. First we train a network with one stripe showing all variations of it, then the second one and then the third one.

Balance calculation (Phase 4)

In the final phase the system calculates how much humans enters and exits the room. Firstly we convert human and movement recognition probabilities to two possible values: 0 or 1. All very high probabilities are converted to 1 and other to 0. Tiny peaks are not used in calculations at all. The direction of movement is calculated using stripes activity information from HTM network.

Results

The proposed schema of the human traffic counting system was tested using 1358 real situations when humans enters and exits a room (see Table 1).

Table 1. Experimental results

Situation	Humans	Accuracy	Direction
One pedestrian	813	83.37 %	76.98 %
Queue	336	77.62 %	73.35 %
Passes each other	209	76.82 %	58.66 %
Total:	1358	80.94 %	73.44 %

All situations are classified into three categories. First category is when one pedestrian walks through passage-way and it is at least few empty frames until another pedestrian walks. Most missed situations are when pedestrians walk very fast or run. The system had the problem with short pedestrians like kids and sometimes one pedestrian is split into two. There are some possibilities to improve accuracy of this category by better training data and by tuning some parameters of algorithm.

Second category is those situations when two or more pedestrians walk in the same direction. The problem with this category is when they walk too close each other.

The third category is when two or more pedestrians walk in the opposite direction and merges in the segment.

An overall accuracy of the system is 80.94 % in the current context.

We have done few comparisons between various kinds of network models by accuracy and by performance. The accuracy difference between TBI and FI modes is shown in Figure 7. Also we compared our 3 levels network versus Visions toolkit 7 levels network (Fig. 8). By performance our network with TBI processes faster 1.6 times than Vision Toolkit network. So, those results show what our network with TBI mode is better tuned for this particular problem.

After using thresholds for recognition results we got front fronts as shown in Fig. 9 and Fig. 10. Humans' quantity now can be calculated by counting end-fronts.

So, using these simple algorithms (making data discrete and analyzing stripes fronts of activity) we can identify if a human enters or exits the room.

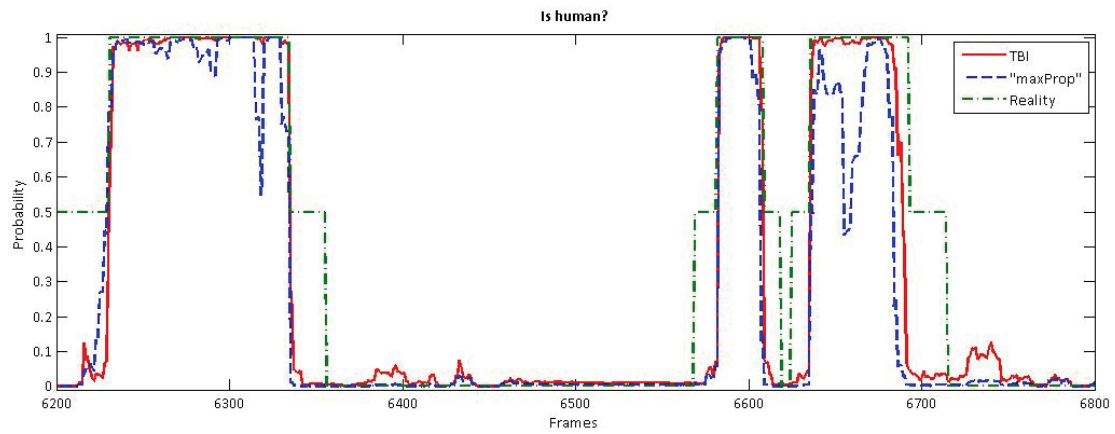


Fig. 7. TBI versus FI (“maxProp”). „Reality“ series visualizes actual values (0.5 – only small parts of humans are visible before entering or leaving, 1 – a human visible in the segment). Certainly TBI makes results more stable and smooth

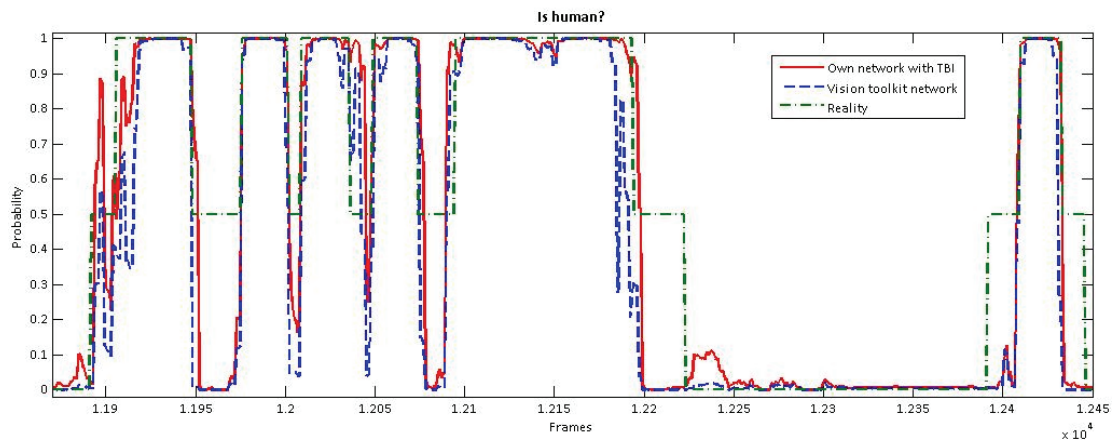


Fig. 8. Our network versus Vision toolkit network. Both networks get similar accuracy results; even with tiny time spans in between

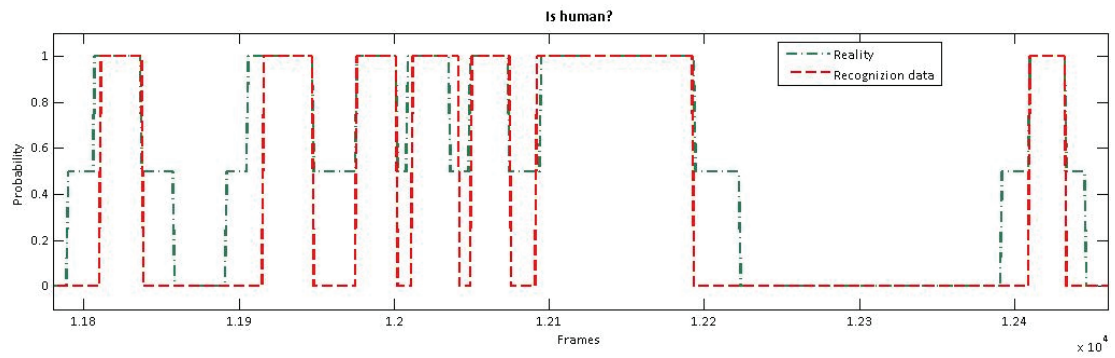


Fig. 9. After humans’ recognition data were processed using thresholds (see Figure 8 to compare with original data)

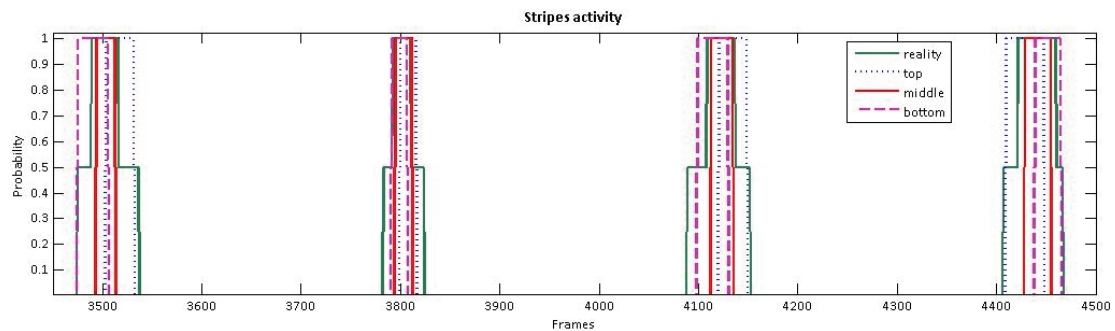


Fig. 10. After stripes activity data were processed using thresholds

Conclusions and future work

In this paper an alternative recognition mechanism was proposed for humans balance calculation in public places using HTM networks. The proposed approach allows distinguishing humans from other objects like background, shadows or small objects in a current video frame, to identify direction of movement, and to calculate how many humans entered or leaved any passageway. Proposed algorithms can be applied to various kinds of traffic balance calculations.

The trained network with three levels is faster and more stable than Vision toolkit optimized networks. With TBI we get better accuracy results but these methods are about 10 times slower, so if there is no necessity to use TBI then FI could be much faster.

The system classified correctly 80.94 % situations of 1358 and accuracy of direction recognition was 73.44 %. As expected, the algorithm works more accurate if just one pedestrian walks.

In the future we will try to use some normalization techniques for input frames. Also we are planning to use infrared camera, because lighting conditions influence on results is essential and we are looking for ways to reduce it.

References

1. **Hawkins J., George D.** Hierarchical Temporal Memory. Concepts, Theory, and Terminology. Online: http://numenta.com/Numenta_HTM_Concepts.pdf.
2. **Chen D., Su C., Zeng Y., Sun S., Lai W., Liao H.** An Online People Counting System for Electronic Advertising Machines // IEEE International Conference on Multimedia and Expo. – New York, 2009. – P. 1262–1265.
3. **Zhu F., Gu J., Yang R., Zhao Z.** Research on Counting Method of Bus passenger Flow Based on Kinematics of Human Body and SVM // International Conference on Computer Science and Software Engineering. – Wuhan, Hubei, 2008. – P. 201–204.
4. **Xin Y., Sun G., Wu Q.** A Preprocessing Method for Tracking and Counting Pedestrians in Bus Video Monitor // International Conference on Industrial Informatics. – Daejeon, 2008. – P. 1689–1693.
5. **Ryan D., Denman S., Fookes C., Sridharan S.** Scene Invariant Crowd Counting for Real-Time Surveillance // 2nd International Conference on Signal Processing and Communication Systems. – Gold Coast, 2008. – P. 1–7.
6. **Ming Y., Wei L.** A New Method for Passenger Flow Counting System Based on Surveillance Video // Second International Conference on Intelligent Networks and Intelligent Systems. – Tianjin, 2009. – P. 453–456.
7. **Jaijing K., Kaewtrakulpong P., Siddhichai S.** Object Detection and Modeling Algorithm for Automatic Visual People Counting System // 6th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology. – Pattaya, Chonburi, 2009. – P. 1062–1065.
8. **Zhao C., Pan Q., Li S.** Real Time People Tracking and Counting in Visual Surveillance // The Sixth World Congress on Intelligent Control and Automation. – Dalian, 2006. – P. 9722–9724.
9. **Teixeira T., Savvides A.** Lightweight People Counting and Localizing in Indoor Space Using Camera Sensor Nodes // First ACM/IEEE International Conference on Distributed Smart Cameras. – Vienna, 2007. – P. 36–43.
10. **Hawkins J., Blakeslee S.** On Intelligence. – Times Books, New York, 2005. – 174 p.
11. **George D. Hawkins J.** A Hierarchical Bayesian Model of Invariant Pattern Recognition in the Visual Cortex // In Proceedings of the International Joint Conference on Neural Networks. – IEEE, 2005. – Vol. 3. – P. 1812–1817.
12. NuPIC 1.7.1. Online: <http://numenta.com/for-developers/software.php>.
13. **George D., Jaros B.** The HTM Learning Algorithms. Online: http://numenta.com/for-developers/education/Numenta_HTM_Learning_Algos.pdf.
14. **George D.** Ph.D Thesis. How the Brain Might Work: a Hierarchical and Temporal Model for Learning and Recognition. – June 2008. – 163 p.
15. **Time Based Inference.** Online: <http://www.numenta.com/for-developers/education/tbi-overview.php>.
16. **Vision Toolkit 1.7.1.** Online: <http://numenta.com/vision/vision-toolkit.php>.

Received 2011 01 05

Accepted after revision 2011 06 07

S. Sinkevicius, R. Simutis, V. Raudonis. Monitoring of Humans Traffic using Hierarchical Temporal Memory Algorithms // Electronics and Electrical Engineering. – Kaunas: Technologija, 2011. – No. 9(115). – P. 91–96.

The main purpose of this paper is to investigate the application of Hierarchical Temporal Memory (HTM) mechanism for humans traffic calculation in a public places using WEB cameras. The proposed approach distinguishes humans from other objects in a current video frame, identifies direction of movement and calculates balance of the humans' traffic. As a result, all people entering and leaving room can be counted and information about people-traffic can be acquired. This information is very useful for further traffic monitoring and can be used for various traffic organization tasks. Ill. 10, bibl. 16, tabl. 1 (in English; abstracts in English and Lithuanian).

S. Sinkevičius, R. Simutis, V. Raudonis. Hierarchinės laikinosios atminties algoritmų taikymas žmonių srauto monitoringo sistemai // Elektronika ir elektrotechnika. – Kaunas: Technologija, 2011. – Nr. 9(115). – P. 91–96.

Pagrindinis šio darbo tikslas yra pritaikyti hierarchinę laikinę atmintį (*Hierarchical Temporal Memory*) žmonių srauto balansui skaičiuoti viešosiose vietose su WEB kamera. Pasiūlyta metodika leidžia atskirti žmones nuo kitų objektų, sekti judėjimo kryptį ir apskaičiuoti srauto balansą. Ši informacija gali būti panaudota optimizuojant sistemas, kuriose žmonių srautai, eilės bei apkrovima yra kritiniai. Il. 10, bibl. 16, lent. 1 (anglų kalba; santraukos anglų ir lietuvių k.).