Satisfaction based Q-learning for integrated lighting and blind control

Zhijin Cheng, Qianchuan Zhao, Fulin Wang, Yi Jiang, Li Xia, Jinlei Ding

1. Introduction

1.1. Background

Building energy consumption accounts for an increasing portion of the total energy consumption. It is found that about 20%–40% of final energy is used for building consumption in developed countries [1]. About 20% of the consumptions come from lighting systems [2], especially in public buildings such as hotels and office buildings. Given more energy saving potential can be obtained via utilizing natural light, it is therefore imperative to develop effective strategies to operate lighting system.

Meanwhile, the comfort requirements are constantly rising [3], it also demands a more precisely controlled luminous environment. In fact, human visual comfort and health are directly affected by the interior lighting level. The survey finds daylight is better accepted due to many people believe daylight is better for psychological comfort, for office appearance and pleasantness [4,5].

1.2. Literature review

1.2.1. Two options of lighting control

In general, there are two options to carry out lighting control: manual or automated control (i.e., fully automated and semi-automated). The users’ behaviors and energy consumptions are different under different control options.

Under manual control, it is observed that the operations of the lights and blinds are usually not adaptive especially when users are busy. People often neglect using the daylight to reduce the electricity consumption for lighting [6]. The lights are seldom switched off even the room is filled with sunshine [6–8]. Similarly, households are accustomed to remaining the blind down once it is manually put down [9].

Under automated control, it is found that lighting control system's energy demand is only about 20% of the manually controlled system [10]. Automated control has been widely accepted by most office workers due to its convenience and energy efficiency. But the users still want to have the authority to override the automated system [11].
1.2.2. Overview of traditional automated lighting control literatures

The existing automated control methods are mainly based on illuminance, transmitted or incident beam solar radiation [10–12] and incident total irradiation or internal temperatures [13,14]. HVAC, lights and shading blinds interact with each other in energy consumption via thermal phenomena and in satisfying human comfort requirements for temperature and illuminance [15,16]. In summer, sunlight is fully utilized for illumination if the blind is up, while some risks including uncomfortable glare [17,12] and more cooling load of HVAC system [13,19,20] are induced with intense daylight. In heating-dominated climates, solar gains are beneficial for HVAC system, where exists no tradeoff between daylight admission and solar heat gain rejection. Thus, the control of the blinds should consider not only the energy consumption of lights but also that of HVAC.

Some integrated approaches, that is, integration of electric lighting control and blind control systems, have been proposed to solve this issue [18,21]. Mukherjee and Birru [22] find the closed-loop integrated controls of blind and lights have higher energy saving potential than the independent control strategies. Additional energy saving can be obtained for the closed-loop control with occupant sensors in private offices [23].

1.2.3. The demands and opportunities for improving automated control

There are opportunities to improve performances of automated control systems by introducing more recent information and machine learning technologies which allow more comprehensive consideration of the balance between user comfort and system energy consumption.

First of all, as we have mentioned, although current automated control has advantages in convenience and energy efficiency, users’ demands are not fully met, so sometimes they still want to override the automated control [11]. The possible reasons are: (1) the existing automated systems are operated according to some recommendations. The unified set-point of illuminance value may not be desired in offices due to the preferred lighting levels varied widely between individuals [8,25,26]. The users’ support will be affected since their personalized requirements are usually ignored [24]. Also, the setting of the blind slat angle still has great room to improve instead of the cut-off angle, yet exhaustively searching for an optimal angle is a challenge due to time-varying solar incident angle [22]. (2) generally, full automated systems are too complex to manage, especially when the system does not run as expected. In the building incorporating automated controlled system, people are usually annoyed and lose their patience since the blinds are perceived to operate at the wrong time [27]. Thus, it may gain more acceptances if there is individual override control provided to users [25,28]. An easy to use human interface is indispensable for both users and facility managers. Users will lose patience to adjust the lighting system if the dimmer is away from their desks [8] [28]. To sum up, personalized control for individuals, intelligent but straightforward automated system with human override control access and friendly interface are essential factors in achieving most user acceptance along with energy saving potential.

Secondly, the energy saving control should consider the coupling among thermal and lighting comfort requirements since indoor temperature is affected by all HVAC, lights and lights and illuminance by lights and shading blinds as shown in Fig. 1.

Third, as an intelligent technique of learning human preferences through human machine interactions, reinforcement learning has been successfully applied in the fields of robotic navigation, helicopter control, as well as intelligent dialogue system [29]. Recently, reinforcement learning has also been introduced into the thermal environment control systems by Dalamagkidis et al. [30]. Considering the subjective property of the visual sensation and individual differences, we believe that there is possibility of using learning method to obtain the personalized preference through interactions. In addition, an improved reinforcement learning algorithm is developed for the challenge of computing an optimal blind angle. Compared with the traditional algorithms, it is event-driven and makes decisions only when users feel uncomfortable, which can reduce the computation significantly. A novel user interface is provided for the users, the inputs of which are straightforward feelings instead of obscure illuminance values. In terms of energy consumption, excessive cooling load of HVAC system is also considered as well as utilizing the daylight to reduce lighting electricity consumption in summer. The presented works focus on the scenario of single-occupancy office. For a group of people sharing the common room, possible extension is discussed in the Section 6.1.1.

2. Brief review of the traditional control strategies

Various lights and blind control strategies have been developed in recent years. Generally, these approaches can be roughly divided into manual control and automated control.

2.1. Manual control strategies

Blind and lights are completely manually operated in this mode. It serves as the baseline for other control strategies when evaluating the energy savings [22]. Users usually directly turned on the lights when arriving at the office and switched off only when leaving in the evening. The blinds were always kept closed in daytime [28].

2.2. Integrated automated control strategies

Mukherjee and Birru [21] introduced prototypes of independent and integrated automated closed-loop control system of lights and blind. It was found that the integrated lighting controls have more energy saving potential than the independent control strategies [22].

This control strategy integrates the lighting and daylight control by sharing the control information (Fig. 2). The main difference between the integrated control and independent control is that the blind system and lighting system independently regulate the light
level to achieve target lighting level while the integrated control operates as one system. To admit more daylight, the integrated controller attempts to exhaustively find a blind slat within the permissible range (i.e., between fully closed and the cut-off angle). Then the lights are incrementally adjusted to compensate for insufficient daylight. In addition, the cooling load of HVAC system is also considered when the room is no occupied or at night in summer, by which the blind slats are completely closed to block solar heat gain [22].

3. The satisfaction based system and learning algorithm

3.1. The satisfaction based system

To improve the performances of the traditional automated lighting control, we propose a closed-loop satisfaction based Q-learning control system (Fig. 3) in this section. It is personalized and human-centric since users’ perceptions of surroundings are employed as the feedback signal. Users’ feelings (e.g., the complaints for glare/dark) are sent to the system via a friendly human machine interface (will be described in Section 3.1.1 and can be seen in Fig. 4). Illuminance is also recorded periodically (i.e., 2 s) by light sensor. Then a comfort model (will be described in Section 3.1.2) will be built based on the feedback signal and illuminance, which could help to judge user’s comfort state. Accordingly, the control policy will be determined by a Q-learning controller (details will be given separately in Section 3.2) and the actuators are operated to adjust the surroundings until users are satisfied. The satisfaction based lighting control can provide a better service customized to user’s requirements, instead of the recommended set-point of illuminance value.

3.1.1. HMI

Well-designed human machine interface (HMI) is an important terms to improve user acceptance [6,28]. Fig. 4 demonstrates a friendly interface employed for the proposed satisfaction based control system (as shown in Fig. 3). Different perceptions icons are designed with different shapes and colors to make them easily distinguish and prevent misuse. Various aspects of interior environment are considered [31], what we focus on here are dark and glare for the luminous environment. The uncomfortable perceptions will be sent to the system after users clicking the perceptions icons of glare or dark. Instead of periodic-sampling, HMI works as an interrupt program which is always ready for the commands once users want to express the feelings. The service of override control is also provided to users in emergency. For example, the button “lamp” and “Blind” on the right of Fig. 4 can directly operate the lights and blind. To facilitate the operation, this HMI can be on a touch panel, a personal smart phone, or press buttons. For personalized control, individual username and password are required when logging the HMI. All these designs would contribute to improving the user acceptance in application.

3.1.2. The comfort model

The comfort model (as shown in Fig. 3) is built to help to judge users’ comfort state for Q-learning. We employ a Pareto frontier set of complaint samples (i.e., perceptions type and illuminance value) to represent the boundary that most probably separated the complaint region and no-complaint region. This design is motivated by the observation that the comfort region is monotonous and bounded in the illuminance when other factors are kept the same, which indicates the boundary description is sufficient to describe the comfort model. The boundary could be described as two threshold values: the lower limit \( M_D \) (threshold value of dark) and the upper limit \( M_G \) (threshold value of glare). Although these values are usually set as constants according to some recommendations, since different people have different preferences, in our comfort model should be adjusted based on the user feedback signal (user

![Fig. 2. Integrated lighting and day-lighting control.](image)

![Fig. 3. The satisfaction based luminous control.](image)

![Fig. 4. The human machine interface.](image)
complaint, i.e., the complaint for dark or glare) and environment data.

3.2. The Q-learning controller

The Q-learning controller in our satisfaction based control system introduced in Section 3.1 (as shown in Fig. 3) is based on a Q-learning algorithm which takes the comfort state information given by the comfort model and sensor information as input and generates command to devices to adjust environment to meet control requirement. In this section, we will first introduce the reinforcement learning framework in which Q-learning is one typical implementation, then present the Q-learning algorithm itself.

3.2.1. The reinforcement learning framework

The objective of the satisfaction based control system introduced in Section 3.1 (as shown in Fig. 3) is to provide a personalized service after learning the individual preference on-line. Reinforcement learning is a general approach to tackle the learning and control problem. A general reinforcement learning problem constitutes the following terms: state variable $s$ (state space $S$), action $a$ (action space $A$), reward function $r(s, a)$, describing the expected reward of choosing a certain action while being in a specific state, value function $V(s)$, representing the total amount of reward that the agent should expect to receive in the long-term by choosing an action while being in a specific state, and policy $L$, representing the way a reinforcement learning agent behaves. It provides a mapping between the situations the agent can find it in (states) and the action it should take. Judging from the above, the reinforcement learning problem becomes the problem of determining the optimal policy (i.e., the policy that achieves maximum rewards over the long run). One of the most common reinforcement learning methods is Q-learning algorithm, which can evaluate the expected utility of the available actions without the transition probability. It directly learns the evaluation of state and action pairs, i.e. Q-factors, to evaluate the value function.

The framework of Q-learning algorithm is shown as Fig. 5. What the Q-learning algorithm concerned with is how an agent should take actions in an environment so as to maximize the cumulative rewards (Q-factor). The learner is not told which actions to take, but instead must discover which actions yield the most Q-factor value by trying them. Based on this framework, what we left is to define the states, actions and rewards for the learning problem. Those definitions need to incorporate the structural information and the properties of the problem to make the method applicable.

3.3. The state variables

Considering dynamic characteristics of the lighting system, the current illuminance in room, the current devices status including the blind status and the lights status are usually chosen as the state variables. Given the inherent characteristics of our system, the subject’s perceptions of surroundings are introduced to substitute illuminance value. Then the state variables are chosen as

$$S = (I, L, B)$$  \hspace{1cm} (1)

where

- $I$ represents the human comfort status, including dark, comfort and glare. In this paper, comfort state is further categorized into the state of saving energy and non-saving energy, denoted as energy saving and comfort state. $I = \{1, 2, 3, 4\}$, which stands for dark state, comfort state, energy saving and glare state, respectively.
- $L$ represents lights status $L = \{0, 1, \ldots, n\}$, which stands for the number of the lights being on, nis total number of the lights in room.
- $B$ is blind status which indicates the blind being down with some certain angle. It is discretized into a series of blind slot angles, $B = \{-20', 0', 20', 40', 60', 80', 90'\}$, where $-20'$ means the blind being up, $0'$ means the blind slot is horizontal, the degrees mean the angle between blind slot and horizontal plan, and the positive degree means the outdoor-oriented edge of blind slot is lower than the indoor-oriented edge. The discrete scale $20'$ is determined by the experience in practice.

The tradeoff between cooling energy and lighting energy is regard as the reward function. The comparison of cooling energy saving with lighting energy is shown in Eq. (2). If the energy saving by putting down blind is larger than the lighting energy consumption, it should put down the blind slat.

$$(P_{in,o}(t) - P_{in,c}(t)) \cdot S_w/COP > P_i(t)$$  \hspace{1cm} (2)

where $P_{in,c}(t)$ is the solar heat gain per square meter from window into room when blind being down and $P_{in,o}(t)$ is the one when blind being up, $P_i(t)$ is lights' total power, $COP$ is the coefficient of performance for cooling [33], $S_w$ is the area of the window. Although the solar heat gain does not transform into cooling load immediately, the solar heat has gone into room and will transform into cooling load in the following time. Therefore for the purpose of simplifying the calculation, here the delay of solar heat transforming into cooling load is neglected and the solar heat gain is assumed to transform
into cooling load immediately. The details are described in the following subsection about reward function. Given the illuminance value detected in time, the next state can be achieved to judge whether the comfort requirements are satisfied, and the actions taken can be decided simultaneously.

3.4. The decision variables

In general, the devices include the lights and blind, and the action $A$ can be

$$A = (\Delta L, \Delta B)$$

(3)

where $\Delta L$ and $\Delta B$ is control command to the lights and the blind, respectively.

The actions of $\Delta L$ may be performed to turn on, hold and turn off one light. Suppose the current light status is $L$, total number of lights in room is $n$, then the next light status $L'$ under the action $\Delta L$ is

$$L' = \min(\max(L + \Delta L, 0), n)$$

(4)

The control of blind has two fundamental aspects to consider, one is its position, including up and down, the other is its angle, from $0^\circ$ to $90^\circ$. Then $\Delta B$ can represent increasing, hold on, and decreasing the angle. Suppose the current blind status is $B$, then the next blind status $B'$ under the action $\Delta B$ is

$$B' = \min(\max(B + \Delta B, -20^\circ), 90^\circ)$$

(5)

where $B' = -20^\circ$ indicates the blind being up.

3.5. The reward function

Reward elucidates the immediate evaluation of the control effects for each action under certain state. Both human comfort and energy conservation should be incorporated.

1) The evaluation of human comfort: With the analysis in the subsection about state variables, the complaints are not only made directly by users, but also generated according to the comfort region $[M_L, M_H]$.

The normalized form of human complaints $I_c$ is

$$I_c = [1, -1, 0]$$

(6)

where “1”, “-1” and “0” represents the glare, dark and others, respectively.

The normalized form of perception estimation $I'_c$ can be denoted by

$$I'_c = \begin{cases} 1, \phi(t) > M_H \\ 0, M_L \leq \phi(t) \leq M_H \\ -1, \phi(t) < M_L \end{cases}$$

(7)

where $\phi(t)$ represents for the interior illuminance at time $t$.

Then the ingredient of the reward for comfort is

$$r_c = -|I_c \cup I'_c|$$

(8)

2) The evaluation of energy conservation: In general, day-light is utilized sufficiently to reduce the electricity consumption when the blind is held up. However, intense solar incidence may induce the risk of excessive cooling load of HVAC system, and then energy conservation is also considered when calculating the reward.

According to Eq. (2), the energy conservation at time $t$ can be defined as

$$I_e(t) = (P_{m,o}(t) - P_{m,e}(t))S_m/(COP - P_l(t))$$

(9)

The calculations of $P_{m,e}(t)$ and $P_{m,o}(t)$ consider the factors including the orientation of the window, the incident angle of sunlight, the blind slat angle, the direct radiation coefficient and scatter radiation coefficient, etc.

Take an east-facing room for example, the details are as follows.

Let $t_0$ be the hour of current time $t$ in the whole year, we can get current horizontal solar radiation corresponding to $t_0$ off line.

$$G(t) = G(t_0) + t_m(G(t_0) - G(t_0 + 1))/60$$

(10)

where $G(t_0)$ and $G(t_0 + 1)$ is the historical data, $t_m$ is the minute in the hour $h$. Then the direct and the scatter part can be calculated as

$$G_d(t) = \mu_d G(t)G_s(t) = \mu_s G(t),$$

(11)

where $\mu_d$ and $\mu_s$ stands for the coefficients of direct radiation and scatter radiation, respectively. With regard to an east-facing room, the radiations in latitudinal direction are

$$G_{d,e}(t) = \varepsilon \cos \theta \cos \theta G_d(t)/ \sin \theta \cos \theta G_{s,e}(t) = 0.5 G_s(t),$$

(12)

where $\theta$ stands for east incident angle and $\theta$ indicates altitude angle, $\varepsilon$ is shading coefficient of building around.

When the blind is up, the thermal power from solar radiation is

$$P_{m,e}(t) = \eta_d G_{d,e}(t) + \eta_s G_{s,e}(t)$$

(13)
While if the blind is down, the thermal power from solar radiation is

\[ P_{m,c}(t) = \eta_{d,c} G_{d,c}(t) + \eta_{s,c} G_{s,c}(t) \]  

(14)

where \( \eta_{d,c} \) and \( \eta_{s,c} \) respectively stands for the transmittance of two layers of glasses for the direct and scatter part with blind up, \( \eta_{d,c} \) and \( \eta_{s,c} \) respectively stands for the one with blind down [35,36].

Based on the analysis and calculation, we can get the energy saving \( P_{m}(t) \) via reducing the thermal load with blind down

\[ P_{m}(t) = (P_{m,0}(t) - P_{m,c}(t)) \frac{S_w}{COP} \]  

(15)

The energy consumption caused by the lights includes two parts, one is the electricity consumption, and the other is the thermal load of HVAC caused by the lighting. We have the following formula

\[ P_l(t) = \omega t_l + \omega n_l / COP \]  

(16)

where \( COP \) is the coefficient of performance of HVAC, \( \omega \) is the power per light, and \( n_l \) is the amount of lights being on.

Thus, the ingredient of the reward for energy is

\[ r_e = L_e(t)/\|k_e\|_\infty \]  

(17)

where \( r_e \) is the normalized form of the reward for energy saving potential.

3) Overall Reward: The overall reward considers both human comfort and energy conservation, and then it is calculated as

\[ R = \rho_c r_c + \rho_e r_e \]  

(18)

where \( \rho_c \) and \( \rho_e \) are weighted factors, which can be adjusted according to actual needs. \( \rho_c + \rho_e = 1, \rho_c \geq 0, \rho_e \geq 0 \), e.g., \( \rho_e = 4\rho_c = 0.8 \) in this paper.

Following are some remarks. First, the perception estimation model will be updated when complaint occurs, while the reward is calculated at each decision epoch. State transition is determined by current states and the chosen actions. However, it is generally random because of uncertainties. In our problem, we focus on the human preference learning instead of the luminous dynamic process. Fortunately, reinforcement learning is especially applicable in this case, i.e. unknown state transition probability.

Corresponding to the reward, the Q factor for the discounted reward case is

\[ Q(s, a) = r(s, a) + \sum_{s' \in S} \gamma P^H(s, s') V^*(s') \]  

(19)

The Q factor is the long-term evaluation for the control actions at each state. Choosing the action that maximizes the Q factors at current state will give optimal control policy. Readers might refer to [29] for more detailed information.

3.6. Several issues and solutions

3.6.1. Target state variables

Large scale state space is the major challenge when applied for practical problems, which makes the standard approaches intractable and even infeasible. Time aggregation techniques [37] are introduced to a substantial reduction in computational and storage requirements. Suppose \( S_1 \) and \( S_2 = S - S_1 \) are two complementary subset of \( S \) mentioned above, where \( S_1 \) is the subset of the states more important or controllable while \( S_2 \) is the part of uncontrolled or negligible. We take actions and update policies only on the important states belonging to these targeted subset \( S_1 \). For the luminous environment control, our concern is when the users feel uncomfortable. Also, the state of energy saving is considered when the users feel comfortable. Then the states involving dark, glare and energy saving are defined as target state variables according to “degrees of importance”.

\[ S_t = \{1, 3, 4\} \]  

(20)

The time-aggregation Q-learning algorithm will take actions and update policies only on the target states, which can obtain a good policy using less computation. It is the reason why the feedback of the luminous perception is required only when the participant feels uncomfortable.

3.6.2. The aggregated reward

Instead of the reward \( r(s, a) \), the aggregated reward \( f(s, a) \) is employed to evaluate the immediate evaluation of the control effects for each action under target state. Naturally, the aggregated reward represents the expected cumulative rewards between two successive target states. Let \( t_i = \min(\{t > t_{i-1}, s_i \in S_1\}, t_0 = 0 \) then we have \( f(s_{t_1}, a_i) = E(\sum_{t_k}^{t_{i+1}-1} \gamma^{t_k} r(s_k, a_k)) \), where \( E(\cdot) \) means the expectation operator. Correspondingly, the Q factor should be adjusted as

\[ \tilde{Q}(s, a) = f(s, a) + \sum_{s' \in S_1} \gamma^t \tilde{p}^s \tilde{V}^*(s') \]  

(21)

where both \( s \) and \( s' \) belong to the target state space, \( \tilde{p} \) is the aggregated transition probability, \( f(s, a) \) is the aggregated reward, \( \gamma \) is the sojourn time between \( s \) and \( s' \), \( \tilde{V}^*(s) \) means the aggregated value function.

3.6.3. Decision epoch

Generally, no special attention was paid on the decision epoch of reinforcement learning and the decision epoch was generally fixed. However, in our problem, it is not applicable because of the stochastic complaint instance and the requirement of real-time response. The decision epoch in our problem is

\[ t_{n+1} = \{\min(t_{c_i}, \tilde{t}_k) | t_{c_i} > t_n, \tilde{t}_k, i, k = 1, 2, \ldots, \tilde{t}_k\} \]  

(22)

where \( t_n \) is the \( n \) th epoch, \( t_{c_i} \) is the \( i \) th complaint instance and \( \tilde{t}_k, k = 1, 2, \ldots, \tilde{t}_k \) is the determinant epoch with fixed interval. The state transition and reward estimation should be done either the complaint occurs or the pre-set time expires. It assures the immediate response of the algorithm when the complaints occur.

3.7. The improved Q-learning algorithm

The details of the proposed algorithm are shown as Algorithm 1.
**Algorithm 1 On-line improved Q-learning Algorithm**

1: Define a target state subset \( S_c \subset S \).
2: Initialize all the aggregated Q-factor \( \bar{Q}(s,a) \) of available state-action for pairs, where \( s \in S_c \), set \( \bar{s}_0 \in S_c, a = a_0 \).
3: Observe the reward \( r(s,a) \) and the next state \( S' \).
4: Judge whether state \( s' \) belonging to the target subset \( S_c \),
    If \( s' \in S_c \),
        then choose \( a' \) from \( A \) using policy derived from \( \bar{Q}(s,a) \) value
        \[ a' = \arg \max Q(s', b) \quad \text{with probability } 1-\epsilon; \]
        \[ a' \in A \quad \text{evenly with probability } \epsilon / |A|; \]
        update Q-factor
        \[ \bar{Q}_{i+1}(s,a) = \bar{Q}_i(s,a) + \alpha_i(f(s,a) + \gamma \max Q(s', b) - \bar{Q}_i(s,a)) \]
    Else
        Do not need to make decisions and update policy.
5: Return to step 3 until \( \bar{Q} \) value is terminal.

According to Algorithm 1, the workflow of the system is depicted as follows.

1) Get current environment data from sensors, i.e., illuminance value in room, the status of the lights and blind.
2) Build a comfort model according to the complaint data, and define the current comfort status of the user (Interpreter module).
3) Determine current state according to the information above.
4) Calculate reward value.
5) Judge any complaint or estimated complaint occurs? If exists, update the Q-factor value and select the action with the maximum-Q factor. If not, maintain the current action.
6) Send the command to the actuator including blinds and light to adjust the environment.
7) Return to step 1).

3.7.1. The convergence of the algorithm

The proposed algorithm is actually a temporal difference method to solve the Bellman optimal equation. Using stochastic approximation theory [38], the convergence can be assured under the same assumption as standard Q-learning algorithm.

4. Experiments and results

4.1. The schematic of the test-bed

To validate our control system design, the satisfaction based control system was implemented in an daily office (6 m × 6 m × 3 m) in Beijing (N37° 48′, E116° 19′). The schematic of the test-bed is illustrated as Fig. 6. It can adjust the various aspects of indoor environment intelligently, where the luminous environment is discussed in this paper. A set of blinds are equipped on an east-facing window (6 m × 2 m) for sun shading and three sets of lights are settled on the roof.

The system can accept the users’ perceptions by a HMI in the form of complaints, simultaneously, the interior illuminance values are collected from the light sensor (settled on the working desk). The technical parameters of light sensor are described in Table 1.

This illuminance value will be collected and stored in system database, which is used to measure the users’ lighting comfort level.

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Range</th>
<th>Accuracy</th>
<th>Output signal</th>
<th>location</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHZD10</td>
<td>0~20000 lx</td>
<td>±4lx/rdg 5dgts</td>
<td>RS485</td>
<td>Aside subject</td>
</tr>
</tbody>
</table>

Table 1. The technical parameters of light sensor.

els. Test-bed snapshots are demonstrated as Fig. 7 and more details are described in [32].

4.2. Experiment settings

The test-bed was implemented in an east-facing room in Tsinghua University in Beijing to support the satisfaction based system and proposed algorithm. Some experiments were conducted from 07:00~11:00 a.m. and 02:00~05:00 p.m. during a day in the summer. 12 candidates with different luminous environmental preferences had participated in the experiments, 5 of whom completed experimental period. One experimental period lasted for about 10 days in different weather roughly categorized as sunny/cloudy. The details of user information are shown in Table 2. Users in the room were asked to keep sedentary and to do easy office work. To avoid excessive disturbance and burden, the participants will express their luminous perception only when they feel discomfort. Some related data such as complaints type, time, user ID and illuminance value have been logged in a database during the experiment process.

Some more requirements are made for users during the whole experiments.

1. People were asked to stay in room in the daytime. Some defined policies may be directly tailored when people leave the office or at night [22].
2. The task users engaged in is required to be easy office work with high contrast and large size (e.g., computer work) during the whole experiment. Some survey found that for people working on computers, the preferred lighting levels were less than the case working without computers [28].

5. Experiment results

5.1. Indoor luminous environment control results

The experiment procedure has two parts: model adjustment and control strategies learning. The comfort model should be adjusted based on the human data and environment data (see Fig. 8(a)). For control strategies learning (see Fig. 8(b)), the comfort state could be judged by the threshold values. The notations of variables could be found in Section 3.2.1, e.g., action vector \( A = (\Delta L, \Delta B) \) (\( \Delta L = \{-1,0,1\} \) represents the action turning on, holding and turning off one light, respectively; \( \Delta B = \{-20,0,20\} \) is the action increasing, holding, and decreasing blind angle, respectively). Suppose initial state \( S = (2,0,20) \) (comfort status is comfortable, light status is on and blind status is up), initial action \( A_1 = (0,0) \), the dark threshold value \( M_L = 200 \) lx, and all Q-factors are 0. At time \( t \), subject inputs a dark complaint using HMI. Meanwhile, if the current illuminance value detected by lighting sensor is 290 lx, then the comfort model is updated as \( M_L = 290 \) lx. Correspondingly, the current comfort state is dark according to the dark threshold value (i.e., \( S = (1,0,20) \)). The action \( A_1 = (0,0) \) results in a dark complaint and gets punishment according to Eq. (18) (i.e., \( R = -0.8 \)). Correspondingly, Q-factor is also updated according to Eq. (21) (i.e., \( Q(S,A_1) = -0.4 \)). Then the action with maximum Q-factor is chosen to adjust the environment. Suppose \( Q(S,A_2) = \max Q(S,A) \), the action \( A_2 = (1,0) \) is chosen (i.e., increase one light being on) to meet user’s requirement. The system state becomes \( S' = (1,1,20) \) (dark is preferred as shown in the last column of Table 2).
Table 2
The users information in the experiment.

<table>
<thead>
<tr>
<th>Experiment location</th>
<th>Experiment date</th>
<th>Number of subjects</th>
<th>Gender (Female/Male)</th>
<th>Age</th>
<th>Career</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tsinghua, Beijing</td>
<td>2012.05.19–2012.10.12</td>
<td>12</td>
<td>5/7</td>
<td>20–40</td>
<td>10 students 2 office workers</td>
</tr>
</tbody>
</table>

Fig. 7. The test-bed snapshots in the experiment.

Fig. 8. The experiment procedure.

Fig. 9 and Fig. 10 demonstrate the luminous environment control using improved Q-learning algorithm for some user under different weather conditions. The glare state is triggered when the subject complains for glare in a sunny day, then a suitable action will be chosen to adjust the environment, e.g., shut the blind with some certain angle. Simultaneously, the room may become dark due to the shade of the blind, then lights will be turned on to provide a comfortable luminous service. The analysis above is described as sub-figure(a) in Fig. 9, where red dots and black solid red dots represent for the complaints for glare and dark, which are made by the subject or estimated by his comfort model, the solid green rectangle markers stand for the energy saving state when the anticipates feel comfortable. The control of the blind (y1 axis of sub-figure(a)) includes the position and closing angle, where the former is being up or down and the latter ranges from 0° to 90° with the fixed 20° intervals. The y2 axis of the sub-figure(a) demonstrates the control of the lights, where the number of the lights being on is optimized during the learning process. The sub-figure(b) demonstrates the illuminance under the control of the presented approach during a day. The situation in a cloudy day is depicted as Fig. 10, from which we can find that blind is always up to maximize the benefits of the daylight. The lights are turned on for auxiliary lighting. The sub-figure(b) indicates that the illuminance under the control satisfies anticipates’ requirements.

5.2. User’s comfort preferences

It is well known that different people have different comfort preferences and requirements. Fig. 11 demonstrates the lighting...
preferences of five users with typical differences during the experiment, where step value represents the dark threshold value, i.e., users input dark complaint at this value. If the step values have no changes for successive 3 days, the preference is regard as being convergent. User3 makes complaints about dark even the interior illuminance is at relative high level (above 400 lx), which indicates he may prefer bright environment. In contrast, less complaints occur for user2 during the experiment when interior illuminance is about 200 lx, which indicates he may have a greater tolerance for dark environment. Moreover, the preferences of all users are convergent for about a week. The case of glare threshold value is similar and needless to say here. Instead of some unified set-point value, these personalized preferences would provide a support when determining the control of the lights and blinds to explore more energy saving potential.

5.3. User acceptances

Both objective indicators and subjective evaluation are considered in the experiment to evaluate the control strategy. Some related data such as complaints type, time, and environment monitoring data have been logged in a database during the experiment process. Also, a questionnaire including the evaluation about the luminous environment control is submitted by the subject after the experiment every day. It is 7-points Likert questionnaire, where score “1” indicates extremely bad while score “7” represents extremely well.

Fig. 12 demonstrates the complaints and evaluation statistics of three typical users during the learning process. From the results in these subfigures, we find the total number of the complaints are reduced overall and hardly any complaints occur in the end. Accordingly, the evaluation scores show a steady rise, which indicates the subjects are satisfied with the control environment in the end. The statistics of users’ acceptances are depicted as Fig. 13. About 92% users vote a relatively high score (≥ 4) and no vote of very dissatisfied score (< 3) exists. Therefore, up to a certain
extent, the improved Q-learning approach can find an optimal policy or sub-optimal policy to provide a comfortable environment for households.

5.4. Energy saving potential

The energy consumptions under manual control [22], traditional integrated control [22] and satisfaction based control are estimated based on simulations under the same days, weather conditions and users. It contains two parts for energy consumptions: consumed for lighting and consumed for cooling caused by lighting and daylight. Table 3 demonstrates the average energy consumptions of different control strategies, where the data of three typical users are chosen for the presented method. The running status of lights and blind (July 23rd, sunny) is depicted as Fig. 14, from which we can find that:

1. Compared with manual control, we find the integrated automatic control strategies can consume less energy, especially under a sunny weather condition. The reason is that more solar incidence can be utilized to reduce the lighting electricity consumption for sunny day. For manual control, energy consumed
Table 3
Energy consumptions of different control strategies.

<table>
<thead>
<tr>
<th></th>
<th>Sunny (kwh)</th>
<th>Cloudy (kwh)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lighting</td>
<td>Cooling</td>
</tr>
<tr>
<td>Manual control</td>
<td>5.18</td>
<td>1.73</td>
</tr>
<tr>
<td>Integrated</td>
<td>3.13</td>
<td>1.52</td>
</tr>
<tr>
<td>Q-learning</td>
<td>Subject1</td>
<td>2.27</td>
</tr>
<tr>
<td>Integrated</td>
<td>Subject2</td>
<td>1.62</td>
</tr>
<tr>
<td>control</td>
<td>Subject3</td>
<td>2.71</td>
</tr>
</tbody>
</table>

Fig. 14. The running status of the devices under three controls (lights & blinds).

for cooling is totally attributed to the thermal load of HVAC caused by the lighting. However, for automated control, the thermal load of HVAC is caused by both lighting and daylight.

2. Compared with traditional integrated automated control, satisfaction based control using improved Q-learning have more energy saving potentials (up to 10%). For traditional control, unification set-point value of interior illuminance would be set at a relative high level (e.g., above 500 lx) since it is difficult to determine what the target illuminance user prefers, which may result in more energy consumption for lighting. In addition, the control of blind angle has great room for optimization where blind slat angle is set to the cut-off angle to block beam solar, which is also responsible for more energy consumption compared to the presented method (Fig. 14). Some integrated controller attempts to open the blind slats incrementally to admit more daylight while ensuring that slat angle stays within the permissible range. It has huge workload to attempt due to the time-varying solar incident angle and users cannot always input their preferences. Actually, the satisfaction based Q-learning method automatically executes the alternative work via machine learning to find an optimal control policy.

3. The energy consumptions of three typical subjects are different under the same presented method (Table 3). It is because that everyone has special comfort requirement and electricity demand, then the presented method can be applied for different people via building a special comfort model for each individual.

6. Discussions

In this section, several issues related to the proposed satisfaction based control method are discussed, including lighting levels with different tasks and areas, the case of multi-users with different preferences, model limitations and possible extensions.

6.1. Lighting levels with different tasks and areas

It was found that the requirements of lighting levels are different with different tasks and areas. In this subsection, possible extensions of the algorithm are explored from simple office tasks to more complex tasks in different areas. Firstly, four lighting levels are categorized with application scenarios based on the Illumination Engineering Society of North America (IESNA) Handbook. The four categories can be roughly defined as low lighting level, medium lighting level, high lighting level and special lighting level. The low level represents the scenario with orientation and simple task. The area could be public spaces, simple orientation for short visits and working spaces for simple visual tasks, e.g., corridors, mess room, storage room and restroom, etc. The activity could be visual tasks only occasional performed, e.g., dining and parking, etc. The medium level represents common visual task with relatively low demands. The area, taking house lighting applications for example, could be living room, lobby and toilet, etc. The activity could be the tasks with high contrast and large size, e.g., easy housework or office work. The high level represents common visual task with relatively high demands. The area could be classroom and kitchen. The activity could be the tasks with high contrast and small size (or low contrast and large size, low contrast and small size), e.g., reading, writing and cooking, etc. The special level represents critical visual tasks. The area could be dentist and sewing room, etc. The activity could be drafting and sewing, etc.

Then the learning processes will be carried out separately with corresponding the category, i.e., at most four sets of comfort regions and control strategies will be learned according to different application scenarios. With the improvement, the presented algorithm could probably be applied but not limited to house lighting applications or office lighting application with complex tasks. More experiments for further verification should be implemented in future work.
6.1.1. Multi-users with different preferences

The key advantage of the proposed control method is its flexibility to provide a personalized service in single-occupancy offices. However, in an actual lighting environment, multiple users are present in a room sharing the same indoor luminous environment and controlled devices. For a group of people, extension is needed for the aforementioned control since comfort zones of group members may or may not be the same. Subjects in our experiment have both bright-preferred and dim-preferred comfort region with 100–300 lx differences. The group dynamic luminous control we would suggest here is a quite natural one: taking the intersection of the individual complaint (dark/glare) regions as the group complaint regions and using the compliment set of the complaint regions as the comfort zone of the group. One benefit of using this approach to define the group comfort zone is its simplicity. Practically, it is also appealing in the sense that it can simplify the modeling procedure in data analysis by simply treating all complaint samples from different subjects as from one subject. Thus, the states in Q-learning controller can be given by the group comfort region instead of the individual comfort region.

Admittedly, the proposed data-driven method is just only from the viewpoint of comfort requirement. The social influence among different users is one of the major factors that affect the group control. One alternative solution for such case is to introduce some group consensus techniques (e.g., game theory). Game theory is a promising technique using mathematical models to handle the conflict and cooperation between decision-makers. The calculation of human comfort and energy savings mentioned in this paper may offer a reference for the formulation of user utility function in the game theory in future work.

6.1.2. Limitation

The present satisfaction based system and Q-learning algorithm also have their limitations. For the system, as one of the automated control system, satisfaction based system can provide a more intelligent service for users while it requires motor-operated blinds and luminous sensors, which may result in more system costs. For the algorithm, as one of machine learning algorithms, Q-learning algorithm can obtain an optimal or sub-optimal control policy after fully interacting with people. To achieve a more accurate comfort model, we need to explore the environment space thoroughly and the subject complained consistently. But the cost of the learning procedure will be dramatic, i.e. it will trigger a lot of complaints and be infeasible in daily office. The current algorithm is a trade-off between the learning cost and model accuracy.

7. Conclusions

In this paper, a lighting system integrating users in the control loop is developed to jointly operate the blinds and lights. Employing an improved Q-learning controller, it can provide a personalized service and improve the indoor luminous environment from three aspects as follows.

(1) Human comfort. As different subjects have different preferences, the presented system could build the individual comfort zones via collecting human data (i.e., complaints) and environment data (i.e., illuminance value). Apparently, a personalized control could provide a more comfortable service.

(2) User acceptance. The evaluation results show that about 92% users vote a relatively high score (≥4) and no vote of very dissatisfied score (<3) exists.

(3) Energy saving. Compared with manual control, the presented algorithm could save energy due to the utilization of daylight instead of lighting. Compared with traditional integrated automated control, the presented control has more energy saving potentials (up to 10%). The recommendations of set-points are usually based on the large average data, while the presented method could discover the threshold value of individual preferences and make an optimal or sub-optimal policy to save more energy. Also, the energy consumptions of different subjects are different since everyone has special comfort requirement.

Future work includes checking the situation of multi-persons in the general place and tasks.

Acknowledgements

This work was supported by the Tsinghua-UTC Research Institute for Integrated Building Energy, Safety and Control Systems, and the United Technologies Research Center. Qianchuan Zhao received additional support from NSFC (Nos 61425027, 61174072, 61174105), the National Science and Technology Support Program 2013BAG18B00 and supported by Innovation Joint Research Center for Cyber-Physical-Society System. This work was also supported by the National 111 International Collaboration Project (B06002) and the TNList Cross-Discipline Foundation. The authors would like to thank the editor and anonymous reviewers for their constructive comments and suggestions that improve the paper significantly.

References


